The role of Gbp2p, Nab2p and Pub1p along the mRNA cycle: structural and molecular recognition studies by NMR and other biophysical techniques

TESIS DOCTORAL

Santiago Martínez Lumbreras

Director

José Manuel Pérez Cañadillas

Instituto de Química Física "Rocasolano"
Grupo de estructura, dinámica e interacciones de proteínas por RMN
Departamento de Química Física Biológica
CSIC

Tutor

José López Carrascosa

Programa de doctorado de Biofísica
Facultad de Química
Universidad Autónoma de Madrid
Madrid, noviembre de 2013

A mis abuelos

AGRADECIMIENTOS

En primer lugar quiero agradecer a José todo su trabajo y dedicación durante estos 5 años. Gracias por enseñarme y compartir conmigo todo lo que he aprendido tanto a nivel científico como personal. Ha sido verdadero honor y una suerte haber podido contar contigo en esta etapa, muchas gracias.

También quiero agradecer a Mª Ángeles su apoyo y labor en la realización de esta tesis. Y hacer extensible los agradecimientos a los demás miembros del departamento: Manolo, Jorge, Marta, Carlos, Doug, David, Miguel Treviño, Cristina, Luis, Irene, Clara y Flor. A su vez, agradecer a las personas que me han ayudado en diferentes etapas de la tesis: Bea con los ensayos de cristalización, Margarita con el dicroísmo circular, Silvia con los experimentos de anisotropía de fluorescencia y Alfredo por los consejos y la ayuda que me ha proporcionado.

I want to thank Bertrand to give me the opportunity to work with him and his team in Strasbourg. It was an enrichment and productive experience. I also want to thank the warm reception the people in his group gave to me, with all those trips, cakes, dinners... it was a great summer.

Y no me podía olvidar de agradecerle a mi gran familia de "Rocalanis". A Yasmina, por ser mi guía y exploradora, abriéndome el camino; a Nerea por haber compartido conmigo todo este proceso (sólo nos ha faltado hacer una lectura de tesis simultánea); a Miguel G. por vivir tan bien y hacérselo vivir a los demás; a Diana porque es un lujo compartir con ella labo, vendimias o problemas de genética; a Soraya y a Miguel M. por "obligarnos" a fumar y a parar para desconectar. También recordar a los que ya acabaron y se marcharon como Miguel P., Jorge Pedro, Fernando, Mar, Inés,… y también a la gente del piso de arriba. Sinceramente, venir a trabajar durante cinco años rodeado así de bien es un placer.

Quiero agradecerles a mis amigos el estar ahí siempre, el saber que nunca me faltará una laurel, unas cañas o una buena cena elegante que compartir, sin importar cuanto tiempo pasó desde la última vez.

A Palma, por toda la compresión y el cariño que me ha dado, sus consejos y su ayuda me han hecho tirar para delante en los momentos más difíciles. Y por último a mis padres y al resto de mi familia ya que todo es gracias a ellos.

RESUMEN

En esta tesis se realiza un amplio estudio biofísico de tres proteínas de unión a ARN de *Sacharomyces cerevisiae* (Gbp2p/Hrb1p, Nab2p y Pub1p) centrándose en su estructura e interacciones con otras biomoléculas. Los resultados se discuten en términos de la función (posible función) de estas proteínas en el ciclo celular del ARN mensajero. Los tres estudios siguen un esquema similar basado en un primer análisis estructural de los elementos de reconocimiento de ARN, seguido de la caracterización de la interacción con ácidos nucleicos y finalizando con un estudio más complejo de la intrincada red de interacciones proteína-proteína en la que participan.

En el primer capítulo, se estudian Gbp2p y Hrb1p, dos proteínas parálogas relacionadas con la transcripción pero cuya función específica es desconocida. Se realiza una extensa caracterización químico-física de diversas construcciones de ambas proteínas observándose diferencias de estabilidad que permiten identificar dos unidades estructurales: el tándem formado por los dominios RRM1 y RRM2, y el dominio C-terminal RRM3. Seguidamente se resuelve la estructura por Resonancia Magnética Nuclear (RMN) de cuatro de sus seis dominios RRM, encontrando motivos estructurales novedosos en los dominios C-terminales de ambas proteínas . Además se realiza un análisis de la afinidad y selectividad por diferentes secuencias de ácidos nucleicos encontrando que de los tres dominios RRM que tiene cada proteína, el primero tiene un reconocimiento canónico de ácidos nucleicos, el segundo presenta una nueva interfase de interacción y el tercero no une ningún tipo de ácido nucleico. La comparación con datos recientemente publicados en la literatura muestran que el dominio RRM2 es similar al dominio RRM2 de la proteína SRSF1 humana y que se conserva el modo de reconocimiento de, al menos dos guaninas consecutivas. En una última sección se investiga la posible función del dominio RRM C-terminal descubriendo que constituye una pieza esencial en el en el reconocimiento del complejo THO, una estructura supramolecular clave para la transcripción y para el transporte núcleo-citoplasmático. Esta importante interacción se ha corroborado mediante el aislamiento y purificación de complejos *in vivo* utilizando la técnica de "Tandem Afínity Purification" (TAP). El análisis de la estructura del dominio RRM3 permite identificar una superficie altamente conservada que involucra al bucle 5 del dominio y encuentra su homólogo estructural en el dominio RRM1 de las proteínas humanas SRSF1 y 9, las cuales también interaccionan con THO/TREX. Estas proteínas tienen una función esencial en "splicing" alternativo, sin embargo, pese a al gran homología con ellas, los experimentos

realizados en esta tesis y los datos existentes en la literatura descartan un papel de Gbp2p/Hrb1p en "splicing" de ARNm.

En el segundo capítulo se realiza un estudio del reconocimiento de ARN de los cuatro primeros motivos tipo dedo de zinc (Zf1-4) de la proteína Nab2p, esencial para el control de la cola de poliadeninas de los ARNm y para su transporte. Se cuantificó la afinidad por diversas sondas de ARNm mediante anisotropía de fluorescencia encontrándose que Nab2p Zf1-4 une con afinidad moderada (K_D~2 μM) oligonucleótidos de poliriboadenosina A_{12}. El estudio de la interacción por RMN no permite extraer conclusiones acerca de la superficie de interacción y por tanto del modo de reconocimiento. Para ello se realiza un diseño racional de mutantes a partir de información estructural de Nab2p y modos de interacción de proteínas similares. Estos estudios permiten identificar una hélice α y el dedo de zinc 3 como las regiones más importantes para el reconocimiento de poli(A) en la región Zf1-4 y junto con información recientemente publicada para la región Zf5-7 por otros autores, proponer un modelo completo de reconocimiento de RNA por parte de Nab2p Zf1-7.

Por último, se analiza la interacción entre las proteínas Pub1p y Tif4631p. Ambas proteínas son homólogas a TIA-1 y eIF4G1 respectivamente, dos proteínas esenciales en el control posttranscripcional en organismos pluricelulares. Al igual que sus homólogas, Pub1p y Tif4631p resultan ser esenciales para el ensamblaje de algunos tipos de gránulos de estrés. En un primer estudio se determina que la región del dominio RRM3 de Pub1p interacciona con la zona N-terminal de Tif4631p. La interacción se corrobora de forma independiente mediante RMN y ensayos de entrecruzamiento. Seguidamente se caracteriza la región N-terminal de Tif4631p, encontrándose que está intrínsecamente desestructurada. La asignación del espectro de RMN de la construcción Tif4631p (1-184) permite realizar titulaciones por RMN con muestras de Pub1p RRM3 no marcadas e identificar las regiones de la primera, potencialmente involucradas en la interacción. Posteriormente un análisis similar utilizando Pub1p RRM3 marcada y péptidos derivados de Tif4631p confirma la interacción de dos de ellos. Uno de estos péptidos coincide un una región altamente conservada en Tif4631p que se demuestra tiene la capacidad de formar de hidrogeles. Además la interacción con Pub1p RRM3 parece modular este proceso de agregación, quizás, semejante al que ocurre en la formación de gránulos de estrés.

CONTENTS

1. Introducción general

1.1. INTRODUCCIÓN

El dogma central de la biología molecular, postulado por Francis Crick en 1958 (Crick, 1970; Crick, 1958), describe el carácter secuencial y direccional del flujo de información desde el ADN a las proteínas (Figura 1.1). Durante los siguientes 40 años, el ARN fue considerado como un simple "mensajero" hasta que se desveló su papel fundamental como elemento regulador de la expresión génica.

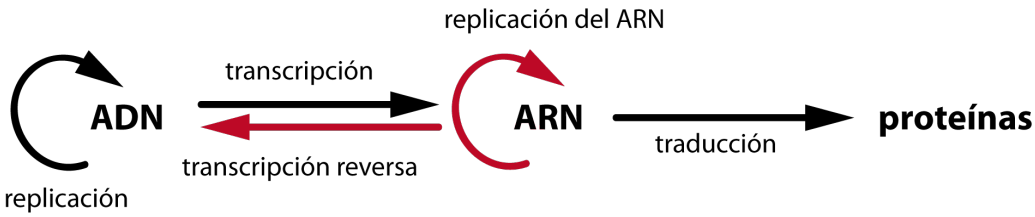


Figura 1.1. – Representación esquemática del dogma central de la biología molecular que ilustra los mecanismos de transmisión y expresión de la herencia genética, incluyendo posteriores modificaciones (en rojo) observadas esencialmente en virus de ARN.

La regulación de los procesos que involucran al ARN se da a diferentes niveles. El núcleo del metabolismo del ARN mensajero (ARNm) es la transcripción, al que se coordinan otros procesos de modificación de ARNm ("splicing", "capping", poliadenilación) (de Almeida and Carmo-Fonseca, 2010; Kaplan, 2013). Además, hoy sabemos que las rutas de degradación de ARNm, que durante años se han considerado como meras estaciones de reciclaje, juegan un papel activo en la regulación (Belostotsky, 2009; Schmid and Jensen, 2008). A su vez, la disponibilidad de ARNm en el sitio y el momento adecuado para traducirse es esencial y por tanto los organismos han desarrollado estrategias de apagado-encendido, que no requieren la producción de nuevos ARNm, denominadas como control post-transcripcional (Belostotsky, 2009)(Figura 1.2).

Por otro lado, la compartimentalización del interior celular tiene una profunda influencia en la complejidad del metabolismo del ARNm. En procariotas, el mensajero sólo se modifica en los extremos 5' (mediante la inclusión de un nucleótido trifosfato) y 3' (gracias a la presencia de bucles o a la poliadenilación) e incluso es traducido antes de que se termine su transcripción. Esto explica la ausencia de grandes rutas de regulación post-transcripcional (Evguenieva-Hackenberg and Klug, 2011). En cambio, en organismos eucariotas los dos procesos de transcripción y traducción están físicamente separados en núcleo y citoplasma.

La polimerasa de ARN II sintetiza el ARN precursor usando como molde una de las hebras del ADN y coordina las diferentes etapas de la transcripción (iniciación, elongación y terminación) principalmente gracias a su peculiar dominio C-terminal, capaz de interaccionar con muchos tipos de proteínas en función de su patrón de fosforilación (Hsin and Manley, 2012).

El proceso de modificación química del ARNm comienza después de la síntesis de los primeros 20 nucleótidos con la adición en el extremo 5' de un nucleótido modificado de guanina (formación del CAP). Esta marca aumenta la estabilidad del transcripto frente a nucleasas 5'-3', pero también juega un papel esencial en la traducción muy posteriormente. La información de un gen se estructura en grupos de secuencias no codificantes (intrones) y codificantes (exones). El proceso de "splicing" permite reconectar estas últimas, pero sobre todo genera una gran variabilidad de combinaciones de exones mediante el "splicing" alternativo lo que potencia la variabilidad genética de forma extraordinaria. Por último, la etapa de terminación transcurre con la formación del extremo 3' (ruptura y poliadenilacion). Una vez completada, la molécula de ARN mensajero maduro se exporta al citoplasma a través de los poros de la membrana nuclear (Figura 1.2) (Bergkessel et al., 2009; Darnell, 2013).

En todo momento, el ARN mensajero se asocia a una gran variedad de proteínas para formar partículas ribonucleoproteicas (RNP o RNPm) con composición dinámica. Por ejemplo, las proteínas encargadas de llevar a cabo las reacciones de "capping" o de "splicing" permanecen puntualmente asociadas, en cambio otras van progresivamente modelando la RNPm (proteínas de unión al CAP en 5', a la cola de poliadeninas,…). Las maquinarias de degradación del ARNm participan activamente en la construcción de las RNPm y su papel como sistema de control de calidad probablemente se relaciona con un ineficiente ensamblado de las partículas (Muller-McNicoll and Neugebauer, 2013).

Las RNPm correctamente formadas se exportan a través de los poros de la membrana nuclear; actividad guiada por interacciones con proteínas de la cara nuclear del poro. Una vez en el citoplasma, las RNPm se remodelan reemplazando proteínas nucleares por citoplasmáticas (p.ej. Pab1p sustituye a Nab2p como proteína de unión a poliadenina) e incorporando otras nuevas que facilitan el reclutamiento de los factores de iniciación de la traducción y las unidades ribosomales (Stewart, 2007). Los propios ribosomas dirigen gran parte de este remodelado durante las primeras rondas de traducción, que sirven de punto de control para evaluar la fidelidad del mensaje (Maquat et al., 2010).

En el citoplasma se establece un nuevo nivel de regulación de la expresión génica en el que la arquitectura de las mRNPs tiene un papel clave en el control de la traducción (regulación post-transcripcional) de determinados ARNm. Partículas competentes para la traducción se pueden inactivar en momentos de estrés (privación de glucosa, deficiencia de aminoácidos, choques térmicos, hipoxia, apoptosis,…), pudiéndose recuperar en condiciones más favorables o por el contrario dirigiendo el ARN mensajero a su degradación en los exosomas (Figura 1.2) (Decker and Parker, 2012).

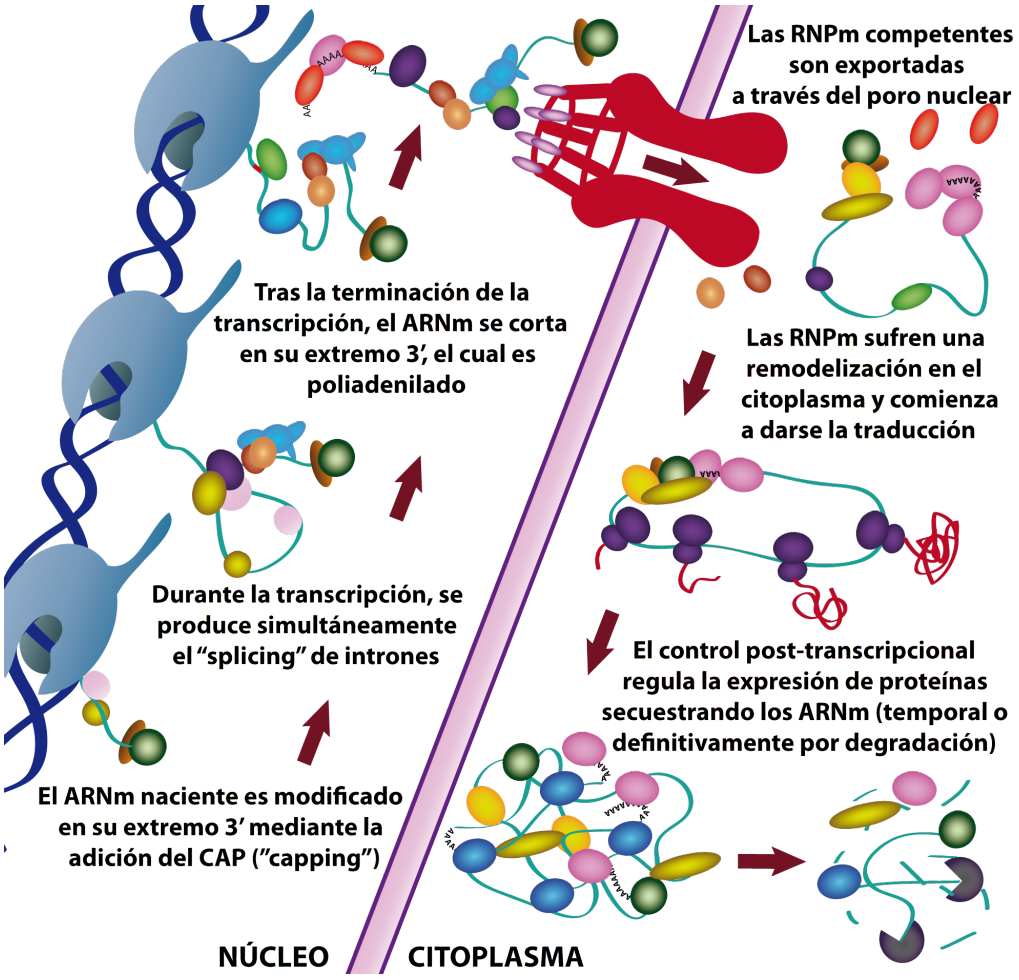


Figura 1.2. – Representación esquemática del metabolismo de los ARN mensajeros en células eucariotas.

Las interacciones entre proteínas y ARNm son esenciales a lo largo de los procesos descritos. Muchas de estas proteínas de unión a ARN se caracterizan por contener: (1) dominios de unión a ARN, (2) sitios de unión a otras proteínas y (3) regiones intrínsecamente desestructuradas.

Existen diversos tipos de dominios de unión a ARN: motivo de reconocimiento de ARN ("RNA Recognition Motif" o RRM), dedos de zinc ("zinc fingers"), de homología K (KH), las cajas RGG y DEAD/DEAH, de tipo Pumilio/FBF (PUF), de unión a doble hebra de ARN (DS-RBD), tipo Piwi/Argonauta/Zwille (PAZ) o los dominios SM (Lunde et al., 2007). Algunos tipos de estos dominios de unión a ARN tienen funciones biológicas muy conservadas y específicas, por ejemplo, los dominios tipo PAZ unen fundamentalmente pequeñas moléculas de ARN de cadena sencilla como micro ARNs o ARN interferentes (Hock and Meister, 2008), o los dominios tipo caja DEAD/DEAH que tienen asociada una actividad helicasa de ARN (de la Cruz et al., 1999). En cambio otras clases de dominios de unión a ARN no tienen a priori ninguna actividad biológica específica sino que participan de forma general en diversas rutas en las que se precisa una interacción con moléculas de ARN (Maris et al., 2005). Lógicamente este tipo de dominios de interacción con RNA son más abundantes (Figura 1.3) y además estos dominios (RRM, Zf y KH) suelen aparecer en más de una copia por proteína (Figura 1.3), pudiéndose encontrar en *S. cerevisiae* proteínas con cuatro dominos RRM, nueve Zf o siete KH (Muto and Yokoyama, 2012).

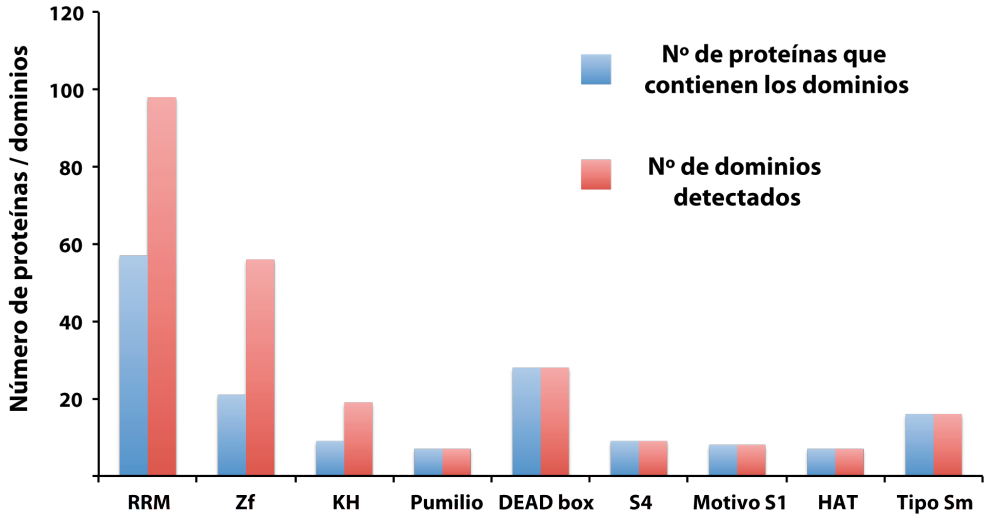


Figura 1.3. – Histograma donde se representan el número de dominios de interacción con el ARN encontrados a lo largo de todas las proteínas de unión a ARN (rojo) y el número de proteínas que los contienen (azul).

Los dominios RRM son los más abundantes entre los motivos de reconocimiento de ARN en levadura y presentan una arquitectura sencilla. Sin embargo, suelen contener multitud de modificaciones que varían su afinidad y selectividad por las moléculas de ARN, incluso llegando a tener otro tipo de interacciones como las proteína-proteína (Daubner et al., 2013). Los RRM son dominios pequeños de unos 90 aminoácidos con una arquitectura tipo $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ en el que las dos hélices se encuentran empaquetadas contra la lámina β antiparalela de cuatro hebras (Figura 1.4). Las hebras β_1 y β_3 contienen dos motivos que dan nombre al dominio: RNP2 ([ILV]-[FY]-[ILV]-X-N-L) y RNP1 ([KR]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY]) respectivamente. En estas dos secuencias se encuentran los tres aminoácidos aromáticos (RNP2 posición 2 y RNP1 posiciones 3 y 5), que refuerzan la interacción con el ARN mediante apilamiento son sus bases nitrogenadas o con los anillos de los azúcares (Clery et al., 2008; Maris et al., 2005).

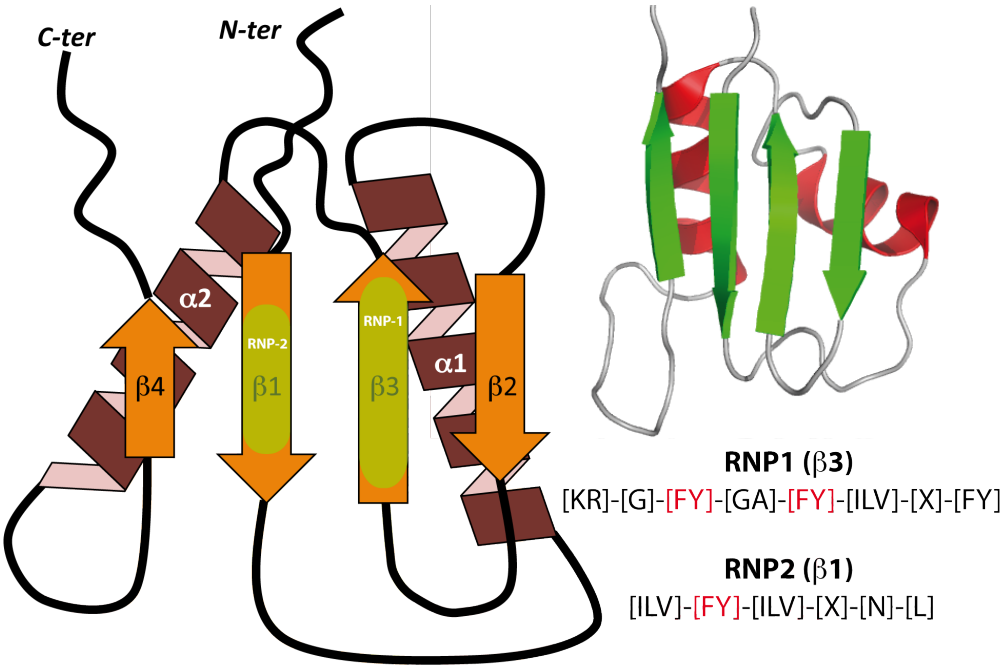


Figura 1.4. – Representación de la estructura canónica de los dominios tipo RRM (izquierda) mostrando las secuencias conservadas RNP1 y RNP2 en las hebras β 3 y 1 respectivamente; ejemplo de la estructura de un dominio (PDB: 2XNQ, correspondiente a el dominio RRM1 de Nab3p).

Sin embargo, se están encontrando cada vez con más frecuencia dominios RRMs con variaciones estructurales sobre este patrón característico: bien porque contienen elementos estructurales adicionales (hélices α, hebras β, bucles) o por variaciones sobre el mismo (p.ej. hélices alargadas, pérdida de un hebra β) (Figura 1.5). Además, se han encontrado modos no canónicos de reconocimiento de ARNs en los que se utilizan interfaces alternativas a la formada por los RNP1 y RNP2. Finalmente cabe mencionar la creciente importancia de los RRM como elementos de interacción proteína-proteína y proteína-péptido. El censo de RRMs a los que se ha adscrito este tipo de interacción es creciente y las superficies implicadas en el reconocimiento (principalmente las hélices 1 y 2) empiezan a conocerse en más detalle. E incluso, se han descrito casos extremos en los que algunos RRM han reemplazado su función de reconocimiento de RNA por el de reconocimiento de proteína (Daubner et al., 2013).

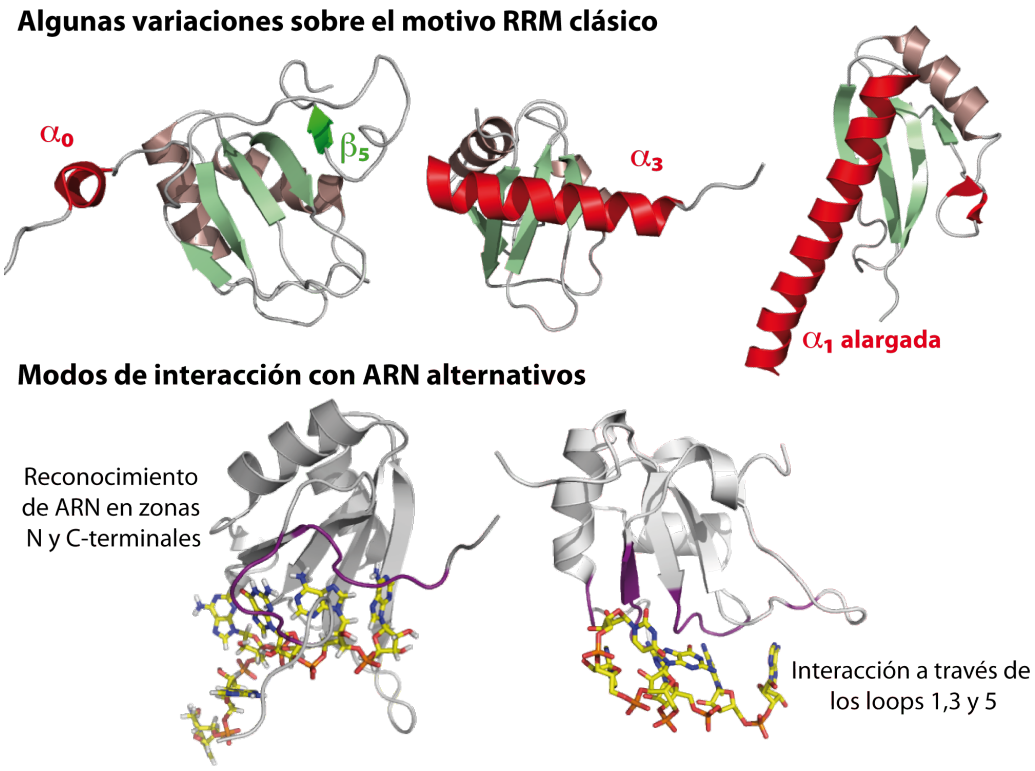


Figura 1.5. – Estructuras de dominios RRM no canónicos con elementos adicionales o alterados y con modos de unión a RNA diferentes (PDBs: 2ADC, 1OWX, 1JMT (arriba), 2RRA y 2KG0 (abajo)).

Los dedos de zinc representan otro grupo de dominios muy versátiles con capacidad de interaccionar con diversas biomoléculas. Los primeros tipos de dedos de zinc se encontraron

en proteínas de unión a ADN, pero actualmente se conocen muchas familias con variabilidad de reconocimiento (ADN, ARN, proteínas, lípidos,…). Todos ellos se caracterizan por ser pequeñas secuencias peptídicas (entorno a unos 25 aminoácidos) que se pliegan alrededor de uno o más átomos de zinc al que se coordinan diferentes cadenas laterales de algunos aminoácidos (fundamentalmente cisteínas e histidinas) (Krishna et al., 2003). Hay varios tipos de dedos de zinc que reconocen moléculas de ARN, los más comunes en levaduras son los de tipo C_2H_2 (con dos cisteínas y dos histidinas en la esfera de coordinación), los CCCH y los CCHC (ambos con tres cisteínas y una histidina) (Hall, 2005) (Figura 1.6).

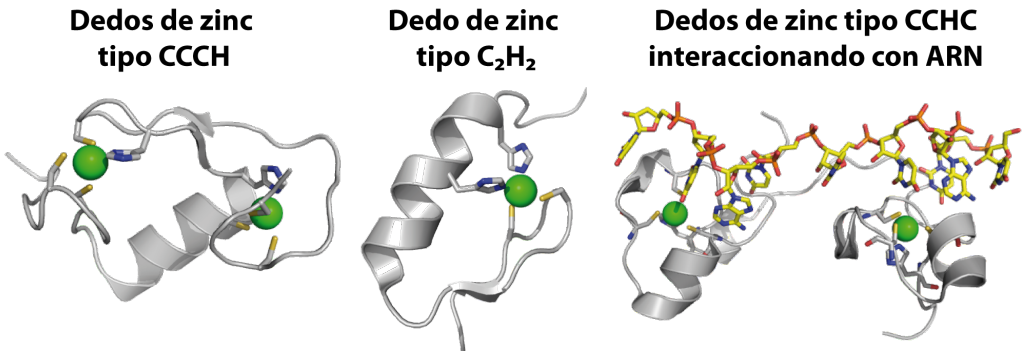


Figura 1.6. – Ejemplo de estructuras de dedos de zinc y su unión con ácidos nucleicos (PDBs: 2LXH, 2EPT, 1RGO).

La versatilidad de interacciones que presentan tanto, los dominios RRM como los dedos de zinc, sitúa a las proteínas de unión a ARN que los contienen como piezas clave para el entendimiento de los mecanismos de regulación de los diferentes procesos del metabolismo del ARNm.

1.2. OBJETIVOS GENERALES

El trabajo realizado en esta tesis está enfocado en diversas proteínas multimodulares que contienen varios dominios de unión a ARN y que participan en la regulación de algunos procesos del metabolismo del ARN mensajero. En todos los casos se ha planteado un estudio interdisciplinar, basado primero en el conocimiento de la estructura tridimensional de los dominios y su caracterización físico-química, para luego proseguir con un estudio de la interacción de los mismos con diversas secuencias de ARN y finalizar con una aproximación más biológica en la que se exploran las diferentes interacciones entre biomoléculas (ARN y proteínas) que aparecen en la etapa correspondiente del metabolismo del ARNm.

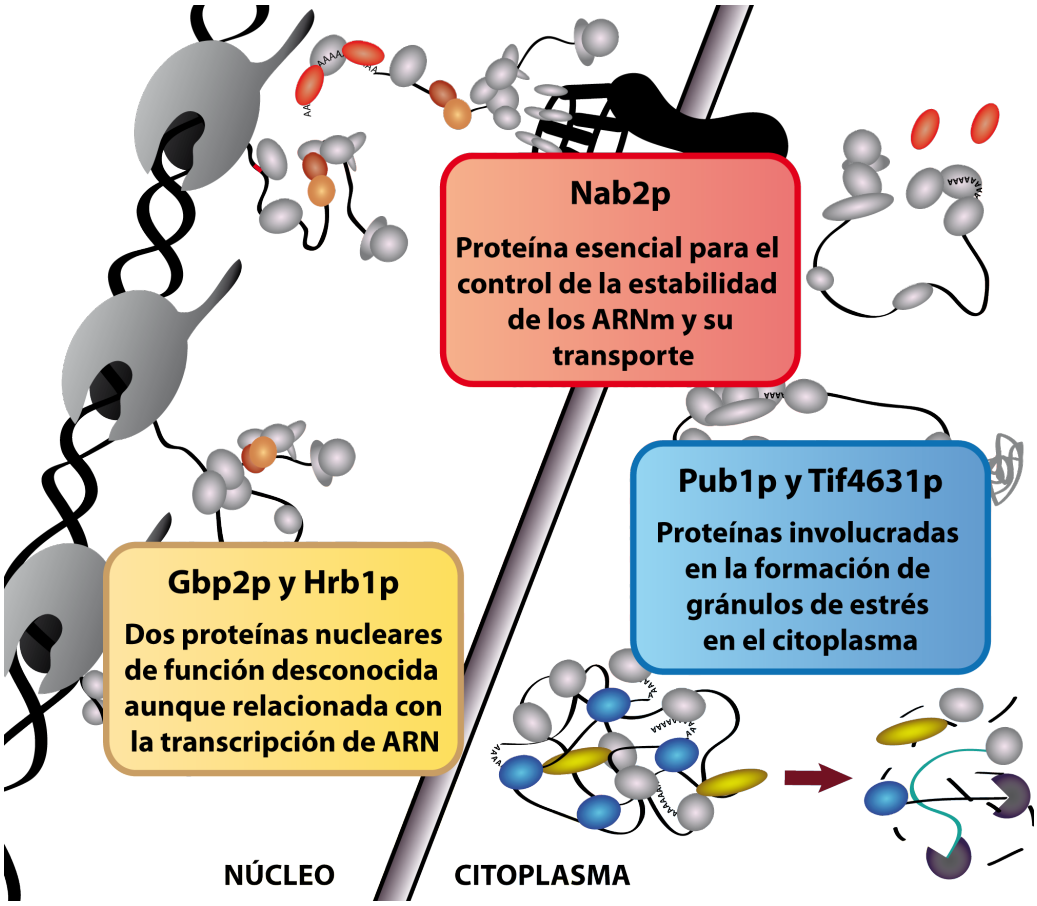


Figura 1.7. – Representación de los tres sistemas estudiados y su función dentro de las diversas etapas del metabolismo del ARN mensajero.

Se han seleccionado tres grupos de proteínas que se encuentran en diferentes etapas del ciclo del ARNm en el organismo eucariota unicelular modelo: *Saccharomyces cerevisiae* (Figura 1.7). En el primer caso se estudian dos proteínas nucleares, Gbp2p y Hrb1p, con función específica desconocida pero relacionada con el proceso de transcripción y transporte al citoplasma de la RNPm. El segundo estudio se ha realizado sobre Nab2p, una proteína nuclear de unión a la cola de poliadeninas en los ARNm, que desempeña funciones en el control de la longitud de la cola y en el paso a través del poro nuclear. Y por último se ha realizado un estudio sobre las interacciones entre proteínas que desencadenan la formación de gránulos de estrés en el citoplasma para el control de la traducción, fundamentalmente el sistema Pub1p-Tif4631p(eIF4G).

1.3. REFERENCIAS

Belostotsky, D. (2009). Exosome complex and pervasive transcription in eukaryotic genomes. Current opinion in cell biology *21*, 352-358.

Bergkessel, M., Wilmes, G.M., and Guthrie, C. (2009). SnapShot: Formation of mRNPs. Cell *136*, 794, 794 e791.

Clery, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. Current opinion in structural biology *18*, 290-298.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crick, F.H. (1958). On protein synthesis. Symposia of the Society for Experimental Biology *12*, 138-163.

Darnell, J.E., Jr. (2013). Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. RNA *19*, 443-460.

Daubner, G.M., Clery, A., and Allain, F.H. (2013). RRM-RNA recognition: NMR or crystallography...and new findings. Current opinion in structural biology *23*, 100-108.

de Almeida, S.F., and Carmo-Fonseca, M. (2010). Cotranscriptional RNA checkpoints. Epigenomics *2*, 449-455.

de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends in biochemical sciences *24*, 192-198.

Decker, C.J., and Parker, R. (2012). P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harbor perspectives in biology *4*, a012286.

Evguenieva-Hackenberg, E., and Klug, G. (2011). New aspects of RNA processing in prokaryotes. Current opinion in microbiology *14*, 587-592.

Hall, T.M. (2005). Multiple modes of RNA recognition by zinc finger proteins. Current opinion in structural biology *15*, 367-373.

Hock, J., and Meister, G. (2008). The Argonaute protein family. Genome biology *9*, 210.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development *26*, 2119-2137.

Kaplan, C.D. (2013). Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae. Biochimica et biophysica acta *1829*, 39-54.

Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003). Structural classification of zinc fingers: survey and summary. Nucleic acids research *31*, 532-550.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nature reviews Molecular cell biology *8*, 479-490.

Maquat, L.E., Tarn, W.Y., and Isken, O. (2010). The pioneer round of translation: features and functions. Cell *142*, 368-374.

Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS journal *272*, 2118-2131.

Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nature reviews Genetics *14*, 275-287.

Muto, Y., and Yokoyama, S. (2012). Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems. Wiley interdisciplinary reviews RNA *3*, 229-246.

Schmid, M., and Jensen, T.H. (2008). The exosome: a multipurpose RNA-decay machine. Trends in biochemical sciences *33*, 501-510.

Stewart, M. (2007). Ratcheting mRNA out of the nucleus. Molecular cell *25*, 327-330.

2. Materials and methods

2.1. MATERIALS

2.1.1 Cell growth media

Two types of cells were used during this work: *Escherichia coli* for bacterial growth and *Saccharomyces cerevisiae* for yeast growth. In both cases, the growth media can be divided in rich and minimal media (also known as chemically defined media). For bacterial growth, Lysogenic Broth (LB) (Bertani, 1951) was used as rich media and K-MOPS (Neidhardt et al., 1974) as minimal media; and for yeast growth, Yeast Peptone Dextrose (YPD) as rich media and Complete Synthetic Medium (CSM) for minimal media (Table 2.1).

Media	Composition
LB	1%(w/v) of tryptone, 0.5% (w/v) of yeast extract and 1% (w/v) of NaCl (autoclaved).
K-MOPS	1:10 dilution of autoclaved K-MOPS 10X solution (0.8 M MOPS pH 8.0 adjusted with KOH, 0.08 M Tricine, 10 mM $MgSO_4$, 10 mM $CaCl_2$, 10 mM $FeCl_2$, 200 µM H_3BO_4, 50 µM $CoCl_2$, 20 µM $CuCl_2$, 100 µM $MnCl_2$, 20 µM $ZnCl_2$, 20 µM Na_2MoO_4, 0.5 M NaCl), 1:1000 dilution of filtered vitamin mix (0.5g/100ml of thiamine, 0.1g/100ml of D-biotin, choline chloride, folic acid, niacinamide, D-pantothenic acid and pyridoxal and 0.01g/100ml of riboflavin), 4 mM potassium phosphate pH 8.0 (filtered), 0.4% (w/v) of $^{12/13}$C-glucose and 0.1% (w/v) of $^{14/15}NH_4Cl$ (depending on the isotopically labelling requirement*).
YPD	1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) glucose (autoclaved).
CSM	0.17% (w/v) yeast nitrogen base (without ammonium sulphate), 0.5% (w/v) ammonium sulphate, corresponding w/v % of CSM mix** and 2% glucose (autoclaved).

Table 2.1. – Composition of the media used in bacterial and yeast growth. * ^{13}C uniformly labelled glucose and $^{15}NH_4Cl$ were acquired from Cambridge Isotopes Laboratories. ** Each commercial CSM mix (*e.g.* CSM-His, CSM-Trp, CSM-Ura) (MP Biomedicals) required different % (w/v), all described by manufacture's chart.

Petri dishes of each type of media were prepared by adding 2% (w/v) of agar to the mix before sterilization in the autoclave. In bacterial media, different antibiotic were used for plasmid selection and were incorporated either to the liquid media or to the plates in the following concentrations: 30 µg/ml of kanamycin (Sigma-Aldrich) or 100 µg/ml of ampicillin (Reig Jofré S.A.). Yeast selection CSM media were prepared eliminating one essential nutrient from the mix (CSM without histidine, CSM without uracil or CSM without tryptophan). Additional antibiotic selection yeast media was prepared in YPD plates supplemented with 300 µg/ml of geneticin (G418, Sigma-Aldrich). In yeast growth tests, glucose in YPD and CSM media was replaced by alternative carbon sources such as: 2% (w/v) galactose or 2% (w/v) lactate plus 2% (w/v) glycerol. For yeast salt stress experiments, plates were prepared adding NaCl to the YPD-agar mix up to 1M final concentration. Finally, other special plates were used

for yeast sporulation, containing 0.25% (w/v) yeast extract, 0.1% (w/v) glucose, 1% potassium acetate and a mix of essential nutrients (adenine, leucine, lysine, methionine, tryptophan and uracil).

2.1.2. Microorganism strains

The two commercial *E. coli* strains were BL21 (DE3) cells (Invitrogen) for protein expression (F-, *ompT, hsdSB* (rB- mB-), *gal, dcm*, (DE3)) and Match1 cells (Invitrogen) for plasmid cloning (F-, Φ80*lac*ZΔM15, Δ*lac*X74, *hsdR*(rK-, mK+), Δ*rec*A1398 *end*A1 *ton*A). *S. cerevisiae* strains obtained in this work were derived of haploid BMA64 (*ade2-1; his3-11,15; leu2-3,112; trpΔ; ura3-1*), courtesy of Dr. Bertrand Séraphin.

2.1.3. DNA plasmids

For protein over expression of Gbp2p, Hrb1p, Pub1p, Nab2p and Tif4631p, constructs were cloned in a home-modified plasmid derived from pET28 (Novagen) in which recombinant mRNA transcription is achieved by the bacteriophage T7 RNA polymerase (Figure 2.1) (Studier and Moffatt, 1986). The system operates by inducing the translation of a chromosomal copy of the T7 RNA polymerase gene, present in strains containing the DE3 genotype, with IPTG (Isopropyl β-D-1-thiogalactopyranoside). This enzyme triggers a massive transcription of mRNAs under the control of T7 promoter (*i.e.* the one present in the plasmid) that are later translated to the fusion protein. The plasmid mRNA contains regulatory elements on its UTRs, like a repressor of T7*lac* promoter, introduced to diminished the basal levels of recombinant protein. Fusion proteins contain an N-terminal thioredoxin A fusion tag, followed by a 6 x histidine peptide, a TEV protease site (ENLYFQ^GS) and the protein construct of interest. Thioredoxin A was used to enhance expression levels, histidine peptide for easy nickel affinity purification and the TEV site was placed to remove this N-terminal portion of the fusion protein. This protease was selected due to its high sequence specificity and the possibility to produce recombinant enzyme at high purity and low prices in the lab. Nab2p constructs and mutants were cloned into a similar plasmid but with glutathione S transferase (GST) as N-terminal fusion protein. This change was made in order to perform an alternative affinity purification step avoiding the use of Hitrap Chelating columns (GE Healthcare) that may capture the zinc ions coordinated to the Nab2p zinc finger motifs.

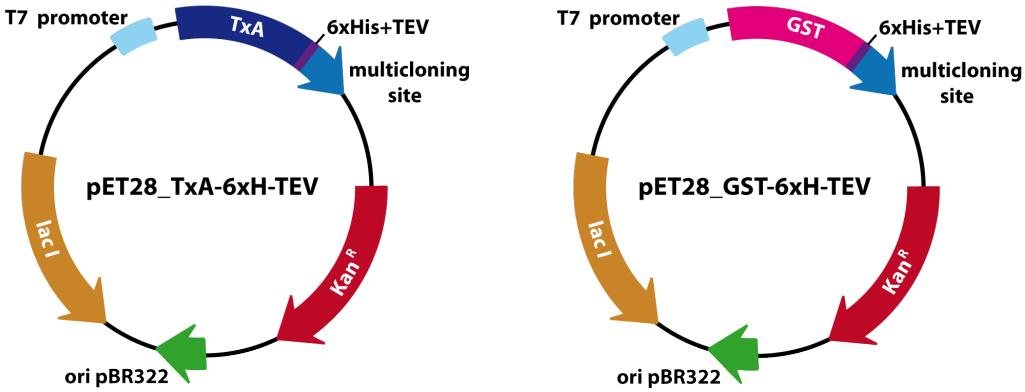


Figure 2.1. – Schematic view of pET28_TxA-HTEV and pET28_GST-HTEV plasmids and their different features.

Many different plasmids were used for β-galactosidase activity tests (Table 2.2). All of them belong to Dr. Bertrand Séraphin's plasmid library (Intitute de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch - France) and are based on the pLGSD5 plasmid (Teem and Rosbash, 1983). These plasmids incorporate the bacterial ampicillin resistance gene and the auxotrophic URA3 gene for yeast selection.

TAP-tag sequences and different selection genes for yeast cloning were obtained by PCR using some plasmids as template. These plasmids were derived from pFA6a ones (Longtine et al., 1998) and also were part of Dr. Bertrand Séraphin's plasmid library (Table 2.2).

Plasmid	Description
pBS-1	Similar to pHZ18, contains the PR51 intron sequence fused to lacZ gen.
pBS-169	Similar to pL6SD5, contains lacZ sequence fused to cytochrome C1 without intron.
pBS-256	Negative control plasmid without lacZ gene.
pBS-979	Derived plasmid of pBS1 with intron mutation that produces pre mRNA leakage.
pBS-52	Derived plasmid of pBS1 with intron mutation in 5' end that produces poor splicing.
pBS-254	Derived plasmid of pBS1 with intron mutation that produces poor splicing.
pBS-255	Derived plasmid of pBS1 with intron mutation that produces pre mRNA leakage.
pBS-983	Plasmid based on pBS1 that contains a synthetic intron sequence.
pBS-2622	pFA6a - TAP T7 Kanamycin.
pBS-2438	pFA6a - TAP T7 Trp1.
pBS-2603	pFA6a - TAP T7 His3.

Table 2.2. – Plasmids used for β-galactosidase activity experiments and for yeast cloning, courtesy of Dr. Bertrand Séraphin.

2.1.4. Synthetic DNA/RNA oligonucleotides

Short nucleic acid oligonucleotides were used in binding studies (Table 2.3) and in PCR amplification of large DNA fragments for cloning purposes (listed in Appendix 1). Oligonucleotides contain standard bases with the exception of those used in fluorescence anisotropy measurements that contain a 3'-end attached fluorescein moiety. All of them were purchased from Integrated DNA Technologies.

DNA Oligo	System	Sequence 5'-3'
TG-43	Gbp2p/Hrb1p	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG
f-TG-43	Gbp2p/Hrb1p	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG-f
YG3	Gbp2p/Hrb1p	TGTGTGGGTGTGTGGGTGTGTGGGTGTG
TG-36	Gbp2p/Hrb1p	GTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG
TG-30	Gbp2p/Hrb1p	GTGTGTGGGTGTGGTGGGTGTGTGGGTGTG
TG-24	Gbp2p/Hrb1p	GGGTGTGGTGGGTGTGTGGGTGTG
TG-16	Gbp2p/Hrb1p	TGGGTGTGTGGGTGTG
TG-12	Gbp2p/Hrb1p	GGGTGTGGTGGG
TG-43m1	Gbp2p/Hrb1p	GTGGTGTGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG
TG-43m2	Gbp2p/Hrb1p	GTGGTGGGTGTGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG
TG-43m3	Gbp2p/Hrb1p	GTGGTGGGTGGGTGTGTGTGTGTGTGGTGGGTGTGTGGGTGTG
TG-43m4	Gbp2p/Hrb1p	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGTGTGTGTGGGTGTG
TG-43m5	Gbp2p/Hrb1p	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGTGTGTG
f-DNA1	Gbp2p/Hrb1p	AAGTAAGTGAGCGCTCACTTACGT-f
f-DNA2	Gbp2p/Hrb1p	CAAAGTAAGTAAATGGTCACTAACGTTGA-f
RNA Oligo	**System**	**Sequence 5'-3'**
SELEX	Gbp2p/Hrb1p	UUGGUGUU
A_{20}	Nab2p	AAAAAAAAAAAAAAAAAAAA
A_{14}	Nab2p	AAAAAAAAAAAAAA
A_{12}	Nab2p Gbp2p/Hrb1p	AAAAAAAAAAAA
A_{10}	Nab2p	AAAAAAAAAA
A_{8}	Nab2p	AAAAAAAA
U_{12}	Nab2p Gbp2p/Hrb1p	UUUUUUUUUUUU
$(AU)_6$	Nab2p	AUAUAUAUAUAU
$(AU)_7$	Gbp2p/Hrb1p	AUAUAUAUAUAUAU
$(CU)_5$	Gbp2p/Hrb1p	CUCUCUCUCU
f-A_{12}	Nab2p	AAAAAAAAAAAA-f
f-A_{10}	Nab2p	AAAAAAAAAA-f
f-U_{12}	Nab2p	UUUUUUUUUUUU-f

Table 2.3. – Inventory of oligonucleotides used in binding studies; -f indicates those nucleic acids with a 6-carboxyfluorescein linked to the 3' phosphate.

2.1.5. Synthetic peptides

The different peptides used in titrations experiments are listed in Table 2.4 and were synthetized by Caslo ApS.

Peptide	System	Tif4631p corresponding sequence	Sequence
Peptide1	Pub1p	37 - 51	GYTNYNNGSNYTQKK
Peptide2	Pub1p	67 - 81	GPNRYNNRGNYNGGG
Box 1	Pub1p	92 - 113	NVPWTGYYNNYPVYYQPQQMAAKKK

Table 2.4. – Inventory of Tif4631p derived peptides used in binding studies with Pub1p RRM3.

2.2. METHODS IN *E. coli* SUBCLONING

The first step of recombinant protein production was the cloning of different plasmids and their transformation into *E. coli* host cells. Plasmid construction was carried out using two different approaches: the cut-and-paste classic method and the site-directed mutagenesis one.

The first and more general method consists on the synthesis of the desired DNA sequences by polymerase chain reaction (PCR), the digestion of their boundaries with restriction enzymes and the ligation of these DNA inserts into linearized plasmids with compatible and dephosphorylated ends. The second method is based on the use of high fidelity DNA polymerases to introduce the desired small changes in the plasmid (*e.g.* punctual mutants or shorter protein versions) and specific DNA nucleases to remove parental (and therefore wild type) plasmids.

Plasmids either prepared by ligation or mutagenesis were first transformed into competent *E. coli* Match1 cells (Invitrogen), selected using the appropriate antibiotic and grown to obtain more plasmid copies. Whatever the method used to generate them, clones were checked by PCR colony screening and confirmed by DNA sequencing. Finally, the plasmids were transformed into competent *E. coli* BL21 (DE3) cells (Invitrogen) for protein expression (Figure 2.2).

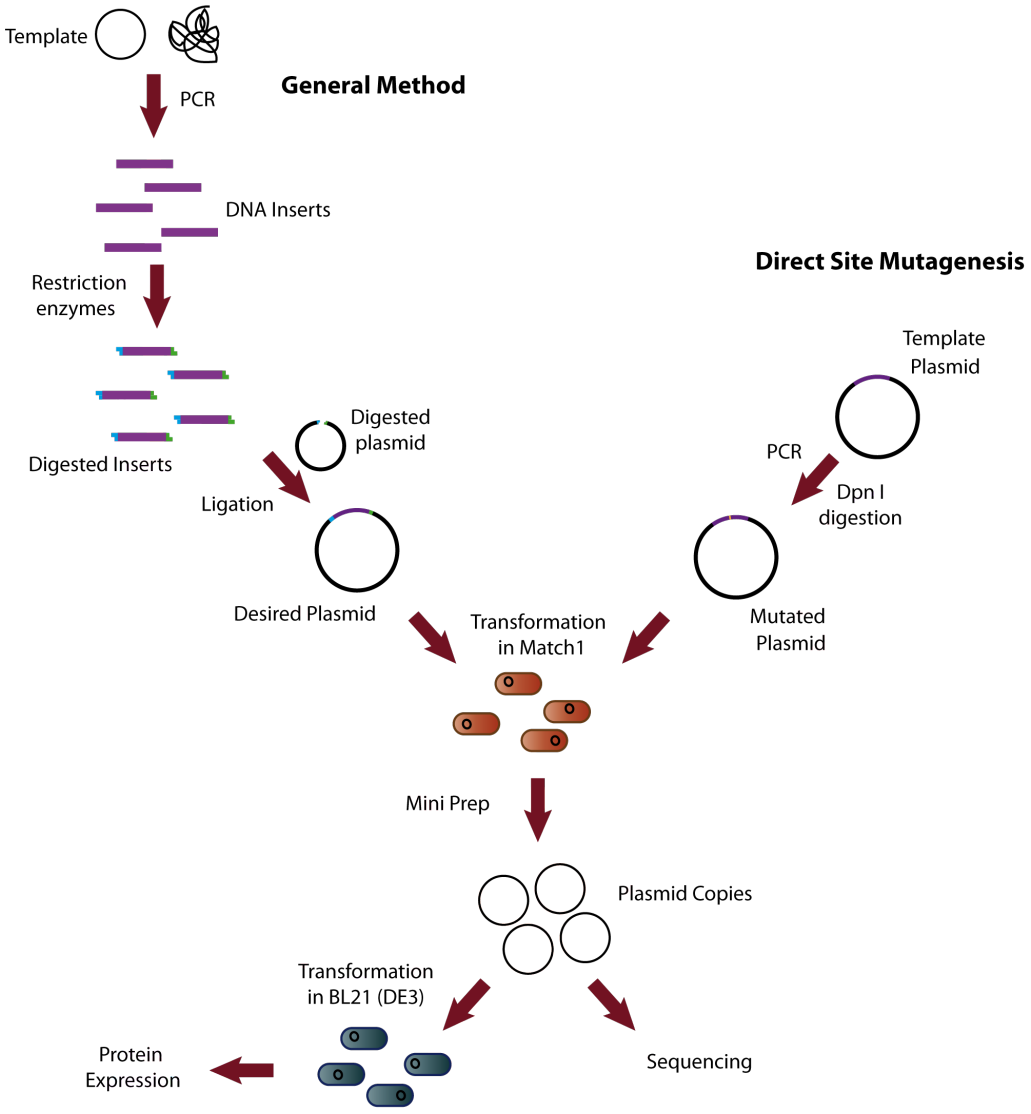


Figure 2.2. – Schematic view of *E. coli* subcloning process.

2.2.1. Polymerase Chain Reaction

DNA fragments for cloning were amplified by PCR using high fidelity DNA polymerase KOD Hot Start (Novagen). For a standard 50 µl reaction, the typical mix contained 1 µl of polymerase (1-2 units), 10-fold dilution of appropriate buffer, 0.5 µM of each DNA primer and the corresponding DNA template (1 µl of purified plasmid or about 10 ng of genomic DNA).

PCR reactions were carried out in a thermocycler (GS-storm) with the following cycle settings: 1 to 2 minutes activation step at 94ºC, followed by 30 cycles of denaturing (15 seconds at 94ºC), hybridization (30 seconds at 58ºC) and elongation (60 seconds per kilobase at 72ºC) steps and ended with an extra elongation step of 10 minutes at 72ºC. Reactions were checked by electrophoresis in 1% - 1.5% (w/v) agarose gels using loading buffer containing a fluorescent DNA dye (Gelstar©, Cambrex).

2.2.2. Purification of DNA fragments

Different DNA fragments were purified following a homemade protocol based on the addition of a chaotropic agent at a 3M final concentration (guanidine chloride or guanidine thiocyanate at pH 8.0), 20% (v/v) of isopropanol and 2.5% (w/v) of diatomaceous earth (Sigma-Aldrich) resuspended in TE buffer. The mix was incubated briefly, separated with a spin column and the DNA, bound to the solid matrix, was washed with room temperature 70% ethanol and eluted with a 20 mM Tris-HCl pH 8.0 buffer.

2.2.3. Restriction digestion of DNA

Both plasmids and inserts were digested using commercial restriction enzymes (TAKARA), typically BamHI, BglII (for 5'-end) and XhoI, SalI (for 3'-end). PCR products were digested with 1 μl of enzyme in the corresponding buffer at 37ºC for one hour, enzymes were deactivated at 70ºC for 10 minutes and DNA was purified as described above. Linearized plasmids were prepared with the same method and dephosphorylated in 5' end (to avoid plasmid recircularization during posterior ligation reactions) using shrimp alkaline phosphatase (ROCHE) following the manufacture's instructions (1 hour at 37ºC).

2.2.4. DNA ligation

Ligations between different DNA inserts and plasmids were performed using a commercial ligation kit (TAKARA). Typically 0.5 μl of digested plasmid (50-100 ng/μl) was mixed with 4.5 μl of insert (≈100 ng/μl) and 5 μl of 2X premix (containing buffer and ligase) and the ligation was carried out overnight at 16ºC. Finally, ligation products were transformed into *E. coli* Match1 competent cells (see protocol below).

2.2.5. Site direct mutagenesis

All mutant protein constructs and some truncation products were generated by introducing changes in a parental plasmid using a site direct mutagenesis kit (QIAGEN). The reaction mix (50 μl) was prepared following manufacture's protocol, using the template plasmid and complementary primers with the desired mutation in the middle of their sequences. The reaction was performed starting with a denaturation step (2 minutes at 95ºC), then 18 cycles of denaturation (20 seconds at 95ºC), hybridization (10 seconds at 60ºC) and elongation (30 seconds per kilobase at 68ºC) steps and finally an elongation step of 5 minutes at 68ºC. All products were treated with 1 μl of DpnI restriction enzyme (Fermentas) to eliminate the template plasmid (methylated) and transformed into *E. coli* Match1 competent cells.

2.2.6. Preparation of *E. coli* competent cells

Match1 and BL21 (DE3) competent cells were prepared following the Inoue method (Inoue et al., 1990). Typically, fresh bacterial cultures were grown in 100 ml of LB media at 20ºC until they reached an OD_{600} of 0.6, then cooled on ice during 10 minutes and finally collected by centrifugation (5 minutes at 3,500 rpm and 4ºC). Cell pellets were carefully resuspended in 32 ml of sterile (by filtration) and cooled Inoue buffer (55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl and 10 mM PIPES at pH 6.7) during 10-15 minutes, pelleted again and resuspended in 8 ml of Inoue buffer with 0.6 ml of DMSO. Cell suspension was incubated 10 minutes on ice, then aliquoted and immediately frozen in liquid nitrogen and stored at -80ºC.

2.2.7. Transformation of *E. coli* with DNA plasmids

For plasmid transformation, frozen Match1 and BL21 (DE3) competent cell aliquots were thawed on ice. Typically, 100 μl of cells were incubated on ice during 20 minutes with 1 μl of purified plasmid or 10 μl of ligation or mutagenesis products. After a heat shock at 42ºC for 45 seconds and 2 minutes on ice, 1 ml of LB media was added and the cells were incubated at 37ºC during one hour. Finally, the cells were pelleted (1 minute at 3,500 rpm), resuspended in 150 μl of fresh media and plated on LB supplemented with suitable antibiotic (100 μg/ml of ampicillin or 30 μg/ml of kanamycin). Plates were incubated at 37ºC overnight (>16 hours).

2.2.8. Minipreparation of DNA plasmids

Plasmid purification was done using a commercial minipreparation kit (QUIAGEN). Typically, cultures (2 ml LB with the appropriate antibiotic) of *E. coli* Match1 cells transformed with the desired plasmid were grown overnight at 37ºC. Cells were harvested (3 minutes at 5,000 rpm), resuspended in a RNAse A containing Tris-HCl buffer at pH 8.0, lysed by the addition of alkaline solution (0.2 M NaOH and 1% SDS) and neutralized using a solution of 3M KAcO at pH 5.0. Mixings were performed gently without vortexing to avoid contamination by chromosomal DNA fragmentation. The cloudy mix was centrifuged at 13,000 rpm for 10 minutes and the supernatant was loaded into the DNA affinity column. Plasmids, bound to the column bead, were washed with a 70% ethanol solution and eluted with a 20 mM Tris-HCl pH 8.0 buffer. The plasmid solutions were stored at -20ºC and their sequences verified in Secugen and Macrogen services using typical T7 and other specific primers. Results were checked using GENtle program (Manske, 2006).

2.3. METHODS IN YEAST CLONING

Different yeast strains used in this work were derived from *S. cerevisiae* BMA64. Gen disruption and protein TAP-tagging strains were performed by homologous recombination of DNA fragments obtained by PCR. To obtain doubled modified strains, the parental haploid cells were matted, the resultant diploids were sporulated and then, the double modified haploid strains were selected. Finally, transformed strains with different plasmids were used in β-galactosidase experiments. These three types of protocols are described in the following sections.

2.3.1. Gen disruption and tagging

For gen disruption and tagging in yeast, typical homologous recombination between DNA fragments obtained by PCR and yeast genome sequence was performed (Puig et al., 1998). First, primers were designed for different cloning strategies, then inserts were produced by PCR and finally these were used for transformation of competent yeast cells. Different procedures were followed to identify and validate positive strains.

2.3.1.1. Primer design

TAP tagging was made at the C-terminus of the target genes. Two oligonucleotides are required for each modification and each one must contain two parts. The 5' half of the forward primers should coincide with the 3'-end of the coding sequence of the target gene (excluding the stop codon), whereas the other half has to match the 5'-end of the TAP-marker cassette. The reverse primers were also designed as hybrids: the 5' sequence has to be complementary to the 3'UTR of the target gene and the 3'-end region should be able to hybridise to the 3'-end of the TAP-marker cassette (Figure 2.3).

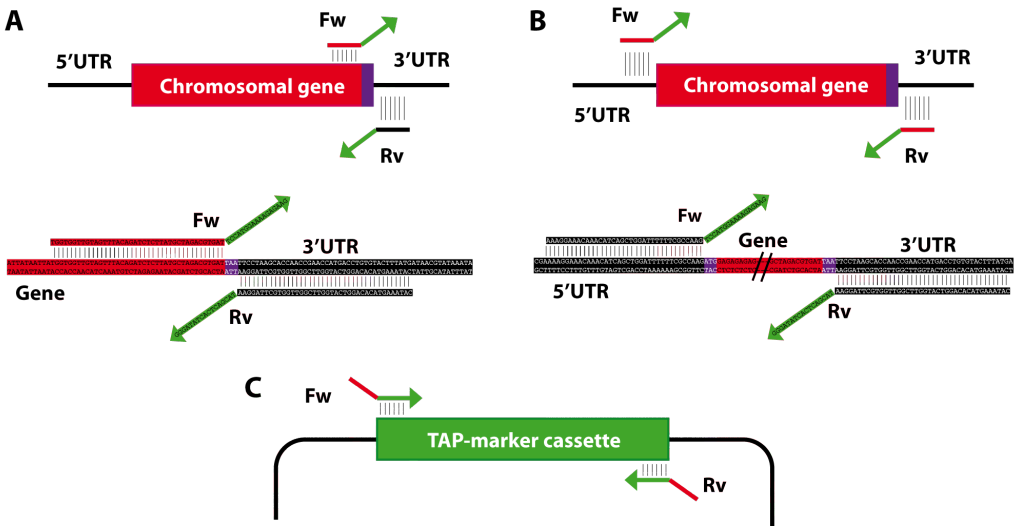


Figure 2.3. – Design of primers for homologous recombination for gen tagging (A) or gen disruption (B). TAP-Marker amplification (C) is common for both processes.

For gen disruption a similar strategy was followed: the homologous sequence of forward primer must contain the end of the 5' UTR and the reverse primer must be homologous with 3' UTR. The 3' end sequence of each primer depends on the type of TAP-marker cassette, as in the TAP tagging (Figure 2.3).

Additional control oligonucleotides were required for PCR checking analysis. They must be complementary to sequences surrounding the modified region in the yeast genome and to the TAP-Mark cassette, as it is shown in Figure 2.4.

2.3.1.2. PCR

DNA inserts for homologous recombination were amplified by PCR using *Taq* DNA polymerase (Takara) and an appropriate plasmid library that contains the desired TAP-marker cassette as template (alternatively, genomic DNAs from strains containing the TAP-marker sequence can be used). For 200 μl of reaction, the typical mix contained 2 μl of polymerase (1-2 units), 10-fold dilution of appropriate buffer, 6 mM of each primer and 4 μl of purified plasmid. PCR reactions started with a 5 minutes activation step at 94ºC, followed by 35 cycles of 30 seconds at 94ºC, 30 seconds at 48ºC and 2 minutes at 72ºC and ended with an extra elongation step of 5 minutes at 72ºC. Reactions were checked by electrophoresis in 1% - 1.5% (w/v) agarose gels and purified following diatomaceous earth protocol (see section 2.2).

2.3.1.3. Yeast transformation

Typically, 50 ml of fresh YPD cultures of wild type yeast were allowed to grow until reaching an OD_{600nm} of 1 and then were harvested by centrifugation (5 minutes at 3500 rpms). Cell pellets were resuspended in 50 ml of 10 mM Tris-HCl buffer at pH 7.5, centrifuged again and incubated for 40 minutes (with generous shaking) at room temperature in 25 ml of a sterile solution containing 10 mM Tris-HCl pH 7.5, 100 mM lithium acetate and 1 mM DTT. Next, the cells were centrifuged again and the isolated pellet resuspended in 0.75 ml of the same lithium buffer. A volume of 100 μl of cells was mixed with 50 μl of lithium buffer, 45 μl of DNA insert (preferably at concentration higher than 100 ng/μl) and 5 μl of sterile carrier DNA (DNA MB grade, Roche). This mix was incubated 10 minutes at room temperature prior to the addition of 300 μl of filtered PEG 4000 solution (2 g PEG in 2 ml of lithium buffer). The mixture was incubated 10 minutes at 25ºC, then 15 minutes at 42ºC, centrifuged at low speed (3500 rmps) for 1 minute and finally, 1 ml of YPD was added to the pellet without resuspension. This mixture was incubated at 30ºC for one hour and then cells were harvested, resuspended in 100 μl of sterile 10 mM Tris-HCl buffer at pH 7.5 and plated on corresponding yeast selective media (CSM-His plates for TAP-His3 based inserts, CSM-Trp for TAP-Trp1 ones and YPD G418 plates for TAP-Kanamycin inserts). Positive colonies should be visible to the naked eye after 2-3 days at 30ºC.

2.3.1.4. Strain screening and quality control

Colonies obtained from transformation were refreshed on corresponding selective plates and a pilot culture of 2 ml of YPD media was grown. The fidelity of the strains was asserted by PCR and western blotting (only for TAP-tagged strains) analyses.

The control oligonucleotides designed in the previous part were required for PCR analysis. First, yeast genomic DNA was extracted from 1 ml of yeast cultures by the following protocol:

Cells were pelleted (1 minute at 14000 rpms) and resuspended in 200 µl of 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM NaCl, 1% SDS and 2% Triton 100x buffer. Approximately 300 µl of glass beads (0.5 mm) and 300 µl of phenol/chloroform /isoamylic acid mix were added and the mix was vortexed for 5 minutes. A volume of 300 µl of water was added and the mixture was centrifuged to isolate the aqueous phase that contains the DNA. This was precipited with 1 ml of iced ethanol and isolated by centrifugation (14000 rpms). Salts were washed away with a 70% ethanol solution and later removed by centrifugation (10 minutes 14000 rpms). Ethanol traces were evaporated by leaving the tubes open to air and the DNA pellet was solubilized in 10 mM Tris-HCl pH 7.5 buffer containing 200 ng/µl of RNAse.

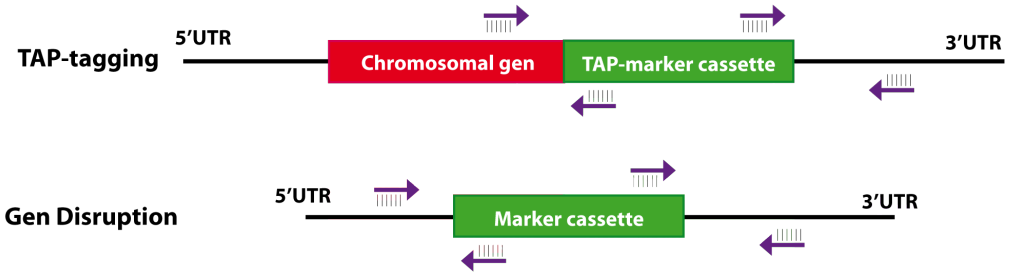


Figure 2.4. – Example of control PCR reactions for correct homologous recombination checking.

An aliquot of 2 µl of a 10 fold dilution of this genomic DNA was used in PCR mix with 12 µM of each primer, 0.5 µl of Taq DNA polymerase (Takara) and the appropriate buffer (schemes of control reactions in Figure 2.4). Reaction was performed starting with an activation step of 5 minutes at 94ºC, then 30 cycles of 30 seconds at 94ºC, 30 seconds at 52ºC and 50 seconds at 72ºC and finally an elongation step of 10 minutes at 72ºC. PCR results were checked by electrophoresis in 1% - 1.5% (w/v) agarose gels.

For TAP-tagged cloning, an additional quality control was performed by western blotting identification of the tagged protein with peroxidase anti peroxidase antibodies (PAP) in each strain protein extract (protein A tags are recognised by this antibody).

The protocol starts from typically 1.5 ml of yeasts cell cultures that were pelleted (1 minute at 14000 rpms) and later resuspended in 100 µl of water. Then, 100 µl of 0.2 M NaOH solution was added and the mixture was incubated for 5 minutes at room temperature. Protein-containing cell debris was centrifuged (2 minutes at 14000 rpms) and resuspended in 100 µl of protein loading buffer, dissolved by heating (3 minutes at 90ºC) and loaded into a polyacrylamide gel. Western blotting was carried out as described on polyacrilamide gel electrophoresis section and a band at the corresponding molecular weight (target protein + TAP tag) should be observed using PAP antibodies (Sigma).

Positive strains according to PCR and western blotting analyses were prepared as glycerol solutions (50%), frozen in liquid nitrogen and stored at -80ºC.

2.3.2. Crossing between two yeast strains

S. cerevisiae can exist in haploid or diploid forms, however only the first type was used for the functional *in vivo* experiments done in the Gbp2p/Hrb1p work. Haploid cells can be sexually differenced in "a" or "α" mating types and crossing could be done between them (a + α = diploid). Additionally, diploid cells can undergo meiosis to produce four haploid cells through sporulation in starving conditions (Figure 2.5).

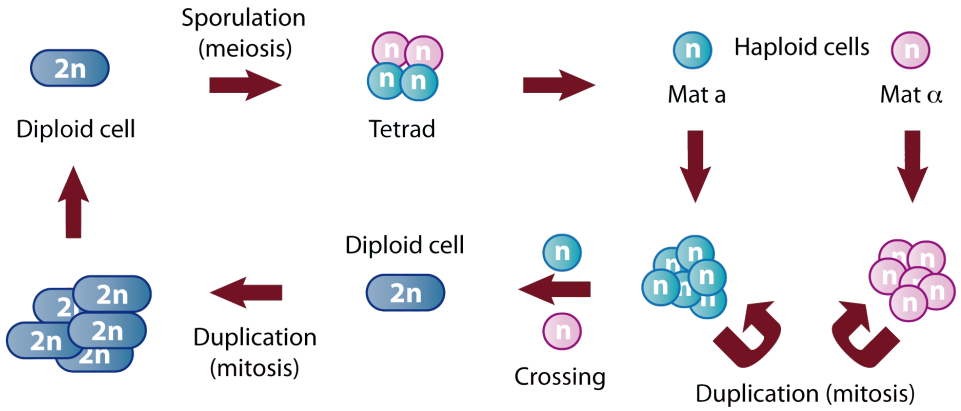


Figure 2.5. – Cell cycle of *Saccharomyces cerevisiae*.

This behaviour is normally used to produce double modified strains by crossing two single mutants. After producing the double mutant diploid cells, sporulation is induced to obtain the double modified haploid strain in one of the four spores (Figure 2.6).

The protocol begins with a matting step, followed by sporulation, tetrad dissection and finally an analysis of the different tetrads in order to find the double mutant one (Figure 2.6). Two important considerations must be taken into account before start crossing: the two strains must have different matting type (one a and the other α) and both should contain different selectable phenotype (*e.g.* tryptophan auxotrophy in one strain and resistance to geneticin in the other).

Matting was performed by mixing both strains in an YPD plate and after overnight growth at 30ºC, diploid cells were selected by sequential platting in agar media containing the two selectable makers (*e.g.* first cells are copied into a CSM-Trp plate and those which had grown were copied into a YPD-G418 one).

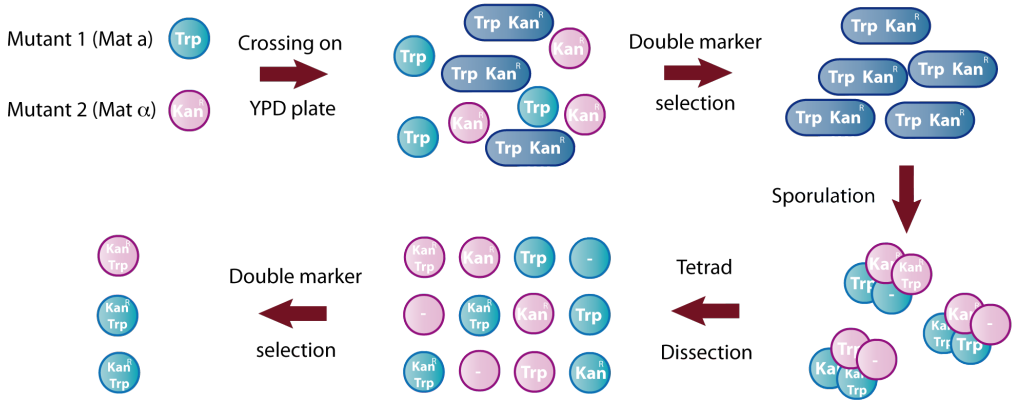


Figure 2.6. – Schematic view of double modified strain preparation by crossing, sporulation and tetrad dissection process.

Sporulation was induced by growing the diploid cells in a plate containing 1% potassium acetate (see 2.1.1 section), during 2-3 days at 30ºC, and was followed by observation under the microscope. Tetrads were dissected when they were clearly developed throughout the plate. Some cells from sporulation plate were resuspended in 50 µl of a buffer containing 1 M sorbitol, 0.1 M sodium citrate at pH 7.0, 60 mM EDTA and a pinch of Zymolyase 100T (Seikagaku) and incubated for 10 minutes at room temperature. The enzyme is used to digest the outer membrane enclosing the tetrad. After incubation, 450 µl of sterile water were

carefully added to the mix, trying to not disturb the cells at the bottom to retain the four spores together. Then, 20 µl of the cells (pipetted from the bottom solution) were plated on a small region of a pre-warmed YPD plate. Using a Singer MSM-400 Dissection microscope, tetrads were localized in the plated region and each spore was placed on a determined quadrant of the same plate using a micro needle. After about 10 tetrad dissections, the plate was incubated at 30ºC during 2-3 days.

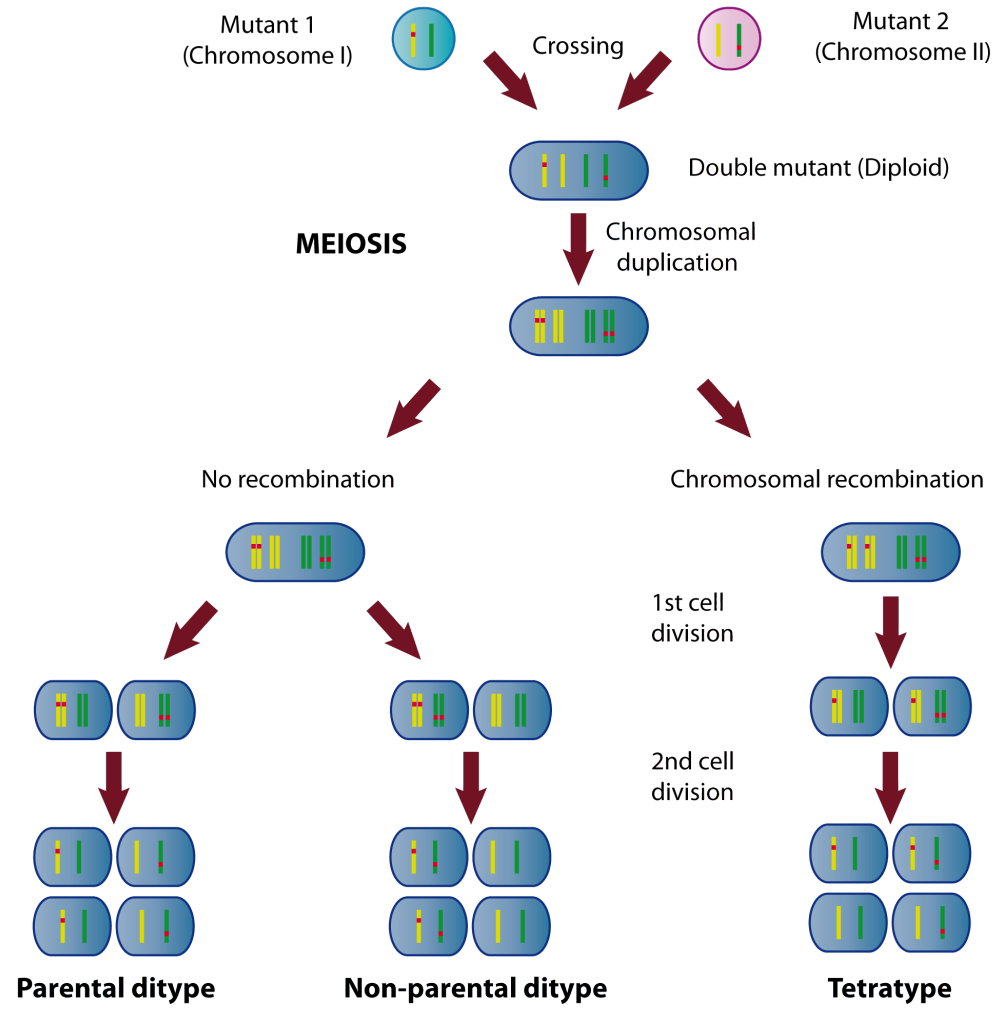


Figure 2.7. – Representation of all types of meiosis for an example of two genomic modifications in two different chromosomes.

Only intact tetrads were analysed in order to find the double mutant strain and each component of the tetrad was grown in selective media for both modifications (*e.g.* in CSM-Trp plates and in YPD G418 ones), identifying the type of meiosis (Figure 2.7, parental ditype, non-parental ditype and tetratype). The double mutant strain was easily detected as the only capable to grow in both media. This procedure was used to obtain the double deletion mutant strain of Gbp2p and Hrb1p and the Gbp2p-TAP tagged strains in ΔHrb1p background.

2.3.3. Yeast transformation with DNA plasmids

The last yeast genetic modification was the transformation of wild type and different mutant strains of Gbp2p and Hrb1p with different plasmids used for β-galactosidase experiments. The protocol consists on mixing 50 μl of a buffer containing 45% PEG 4000, 0.1 M lithium acetate and 100 μg/ml of carrier DNA, with 1 μl of the desired plasmid (about 100 ng/μl) and a pinch of fresh yeast cell strains (from a YPD plate). This mixture was incubated for 50 minutes at 42ºC and then pelleted, resuspended in 100 μl of sterile 10 mM Tris-HCl buffer at pH 7.5 and plated on corresponding selective media (histidine auxotrophy gen HIS3 is included in all used plasmids). Plates were incubated at 30ºC for 2-3 days to isolate the colonies that were used straight away in different experiments. In contrast to yeast genomic modifications, yeast strains transformed with plasmids require chemically defined growth media with auxotrophy selection in order to maintain plasmid (as it happens in bacterial transformation). Here, the transformed Gbp2p/Hrb1p strains were grown in CSM-His.

2.4. METHODS IN BACTERIAL PROTEIN EXPRESSION AND PURIFICATION

2.4.1. Protein expression in bacterial system

All protein samples were overexpressed from plasmids based on bacterial T7 expression system, where the expression of the target protein is regulated under T7 RNA polymerase bacteriophage promoter. *E. coli* BL21 (DE3) cells contain a chromosomal copy of the T7 RNA polymerase gene under the control of the *Lac*UV5 promoter. The addition of IPTG activates the transcription of this gene, which triggers high-level transcription of the target gen in the plasmid, as it is controlled by the T7 promoter. This causes a hyper accumulation of the heterologous mRNA that is finally used by the bacterial machinery to produce large amounts of the target protein.

Typical expression procedures consisted on the inoculation of 500 ml of rich (LB) or minimal (K-MOPS) media with a saturated culture (500 μl) of *E. coli* BL21 cells transformed with the appropriate plasmid. Bacterial cultures were grown at 37ºC until reaching an OD_{600} of 0.6; then, they were cooled at 20ºC and when temperature was stable, IPTG was added at a final concentration of 0.5 mM for induction. After overnight (> 16 hours) growth at these low temperature conditions, cells were harvested (15 minutes at 3500 rpms) and either processed straightaway or frozen with liquid nitrogen and stored at -20ºC. For Nab2p protein expression, an addition of $ZnCl_2$, at a final concentration of 10 μM is required prior to induction step with IPTG (typically when culture reaches an OD_{600} of 0.3).

2.4.2. Protein solubility and expression tests

Small-scale protein expression tests were performed in order to evaluate the level and quality of the recombinant protein expression and if it is suitable for scaling-up. Typically, 5 ml of LB (with the required antibiotic) were inoculated with a bacterial culture, transformed with the appropriate plasmid, and grown until reaching an OD_{600nm} of 0.6. Then, they were induced with IPTG (final concentration of 0.5 mM) at the desired temperature (37ºC or 20ºC) and harvested after 3 hours at 37ºC or overnight at 20ºC. About 100 μl of cell culture were isolated (1 minute at 14000 rpm) and resuspended in 50 μl of a detergent mix (BugBuster, Novagen). A 10 μL sample of this solution was taken at this point (total fraction) and the rest was centrifuged during 10 minutes at 14000 rpm to collect a second sample of the supernant (soluble fraction). Both samples were analysed by SDS-PAGE to determine levels of expression and solubility.

2.4.3. Cells lysis

Cell pellets from large-scale cultures were resuspended in ice-cooled lysis buffer (30-40 ml for a 500 ml culture pellet) containing 20 mM potassium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 1% β-mercaptoethanol, a tablet of protease inhibitors (complete mini EDTA-free, Roche) and ~12 mg of lysozyme. With the exception of cells containing overexpressed Nab2p constructs that were resuspended in PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4 at pH 7.4) with a tablet of protease inhibitors. Preparative cultures of Gbp2p, Hrb1p and Nab2p were typically processed by sonication whereas Pub1p and Tif4631p ones were lysated by French press.

The sonication protocol used an Ultrasonic Processor 750W (Sonics) at 40% of power amplitude applied on 5/20 seconds on/off cycles for 3-4 minutes of total sonication time. To avoid protein damage during the process the cell suspension was kept overcooled using a salt-ice bath. Lysis protocol with French press (THERMO) was performed at a pressure of 12000 psi, and each lysate was passed twice through the press. To clarify lysates an ultracentrifugation step was carried out for 30 minutes at 30000 rpm in an ultracentrifuge (Centrikon T-1075).

2.4.4. Protein purification protocol

All protein constructs of Gbp2p, Hrb1p, Pub1p and Tif4631p were purified by a common chromatography protocol (Figure 2.8.A) that consists on the following steps:

- Nickel affinity chromatography to capture recombinant proteins through their 6 histidines tag.
- Digestion with TEV protease to separate the target protein from the TxA-6xH tag.
- A second nickel affinity chromatography step to capture all contaminants (non-digested proteins, TxA-6xH tags and TEV protease which includes a 6xHis tag).
- A final ion exchange chromatography (anionic or cationic) to concentrate and purify the proteins.
- Concentration and exchange to the final buffer.

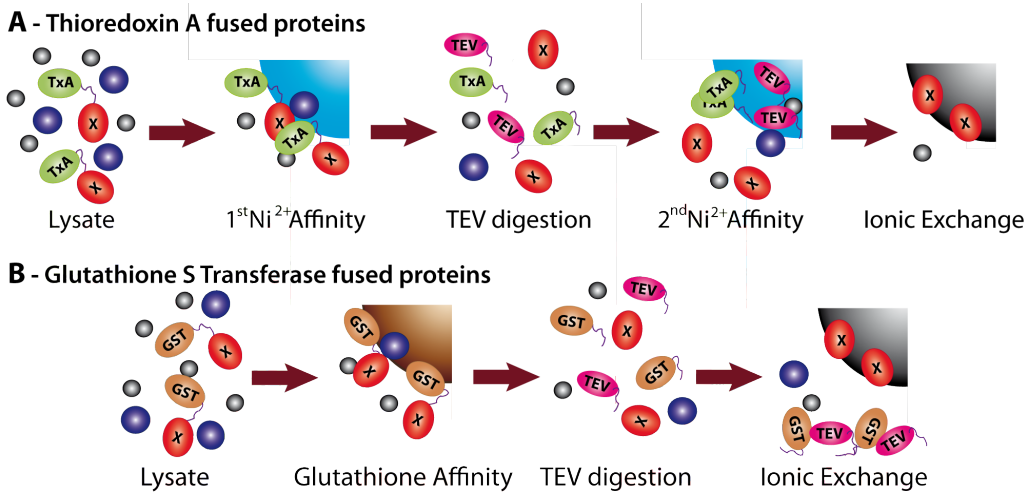


Figure 2.8. – Graphical overviews of the protein purification process followed for thioredoxin A (A) and GST (B) fused proteins.

Nab2p constructs and mutants were purified following a similar protocol, but replacing the nickel affinity chromatography for glutathione affinity one (Figure 2.8.B). In addition, as GST digested protein and TEV protease bind to a different ion exchange column than Nab2p constructs, no second affinity chromatography step is required after TEV digestion.

2.4.4.1. Nickel affinity purification

Clear lysates obtained after extract centrifugation were loaded into a 5 ml nickel Hitrap Chelating column (GE Healthcare) previously equilibrated (20 mM potassium phosphate pH 8.0, 300 ml NaCl, 10 mM imidazole and 1% β-mercaptoethanol). Weakly bound proteins that do not contain the 6xHis tag were washed away with a buffer containing 20 mM potassium phosphate pH 8.0, 300 ml NaCl, 20 mM imidazole and 1% β-mercaptoethanol. The pure recombinant protein was eluted with a high concentration imidazole buffer (20 mM potassium phosphate pH 8.0, 300 ml NaCl, 300 mM imidazole, 1% β-mercaptoethanol). All processes were monitored in FPLC and the elution peak was recovered for next step and checked by SDS-PAGE.

Digestion with TEV protease was carried out using homemade enzyme (100-200 μg/ml) during an overnight dialysis at 4ºC against a 20 mM potassium phosphate pH 8.0, 25 mM NaCl and 1% β-mercaptoethanol buffer. The extension of the proteolytic reaction was checked by SDS-PAGE until completion (usually < 16h at 4ºC) prior the reverse nickel affinity step. A small number of fusion prodcuts were resistant to TEV digestion and required further aliquots of TEV and/or performing the enzymatic digestion at higher temperatures (16-20ºC). TEV protease diminish is enzymatic activity over time (< 12 hours), thus when more than one aliquot was required, it probed to be more efficient to add them stepwise rather than all in a single step.

Digestion products were loaded into the same 5 ml nickel Hitrap Chelating column used in the first chromatography step (equilibrated with 20 mM potassium phosphate pH 8.0, 300 ml NaCl, 10 mM imidazole and 1% β-mercaptoethanol buffer). Flow through was collected until monitorized UV comes down to the level of the buffer, then the column was washed with buffer containing 20 mM imidazole and finally bound proteins were eluted with 300 mM imizadol containing buffer. All the samples (flow through, wash 1, wash 2 and elution) were checked by SDS-PAGE and those solutions containing the target protein (normally flow through) were pooled together to proceed with the next purification step.

2.4.4.2. Glutathione affinity purification

Soluble lysates of Nab2p constructs were mixed with 1/5 volume (typically 10 ml) of Glutathione Sepharose 4 Fast Flow (GE Healthcare) and rocked on ice for 1 hour. Unbound fraction was filtered out in a gravity column, the resin was washed with about 50 ml of PBS and fusion proteins were eluted with a buffer containing 50 mM Tris-HCl and 10 mM reduced glutathione at pH 8.0. Elutes were checked by SDS-PAGE and dialyzed against a 20 mM Tris-HCl buffer at pH 8.0. TEV digestion treatment was performed following the same protocol used for thioredoxin A fused proteins.

2.4.4.3. Ion exchange chromatography

Because this type of chromatography requires loading the proteins at low ionic strength, the buffer must be exchanged to remove the high salt concentration. Two procedures were followed depending on the stability of the protein, a dialysis step (Gbp2p, Hrb1p and Pub1p) or a buffer exchange chromatography (Tif4631p). In contrast, Nab2p proteins could be purified directly without dialysis as they were already in a low ionic strength buffer after TEV digestion (20 mM Tris-HCl at pH 8.0).

Protein products were loaded into an ion exchange column (5 ml Hitrap SP or Q, GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris-HCl pH 8.0, 25 mM NaCl and 1 mM DTT. Preliminary experiments were done to determine if the target protein is an anion (Q) or a cation (SP) binder. All constructs from Gbp2p, Hrb1p and Tif4631p proteins and Pub1p RRM12 construct bound to the anion exchange column (Q) whereas Pub1p RRM3 bound to the cation exchange column (SP). The proteins were first loaded into their respective columns and then eluted applying a linear salt gradient (from 25 mM to 1 M NaCl) in a buffer containing 20 mM Tris-HCl and 1 mM DTT. In the case of Nab2p constructs, TEV digestion products were loaded into both SP and Q columns that were arranged in series in a specific order (SP-Q). GST tag and TEV protease bound to the cation exchange column (SP) and Nab2p constructs (and mutants) to the anion exchange one (Q). Then, columns were separated and eluted using the same linear salt gradient described before. The exception was Nab2p Zf3-4 that did not bind to any of the two columns and came out in the flow through.

Different fraction samples were monitored by UV-absorbance in FPLC and also by SDS-PAGE. The samples containing the desired protein in a high purity level were pooled together.

Depending on the final application requirements, proteins could be concentrated and/or dialysed or stored at -20ºC.

2.4.4.4. Protocol modifications for intrinsically unstructured proteins

Well-folded protein constructs of Gbp2p, Hrb1p, Nab2p and Pub1p are resistant to unspecific enzymatic and chemical degradation. However this is not the case of intrinsically unstructured constructs of Tif4631p. These proteins require special adaptations of the protocols described above in order to guarantee sample quality and integrity. In general the purification steps have to be performed as quickly as possible, often replacing long dialysis steps by fast chromatography buffer exchange methods.

The use of nickel columns causes inevitable low levels of metal contaminants (leakage) that co-elute with the purified proteins. These metals could catalyse chemical modification and degradation of the recombinant proteins. Hence, in the nickel chromatography steps of Tif4631p constructs, Tris-HCl-based buffers were used instead of phosphate-based ones, because the former exhibit superior protein stability, probably due to weak metal-chelator effects. To potentiate this effect, up to 100 μM of EDTA was added to eluates of Ni^{2+} columns. In addition, to remove any remaining traces of metal or metal chelates, these fractions were buffer exchanged with a preparative column, and not dialysed, immediately after collection from the nickel column.

2.4.5. Buffer exchange methods

Buffer exchange was required at some steps of the purification protocols as well as to prepare the protein samples for the final applications. Typically, during purification all buffer exchange procedures were made using dialysis (large volumes), except for Tif4631p constructs (see above for special directions). Once proteins were purified and concentrated, buffer exchange could be done either by microdialysis or using commercially available desalting columns.

2.4.5.1. Membrane dialysis

All protein samples were dialyzed overnight at 4ºC and using membranes with a molecular weight cut off (MWCO) of 3500 Da (Spectra/Por, Spectrum Lab). Typically, the minimal ratio between sample and dialysis buffer volumes was 1:50.

2.4.5.2. Buffer exchange chromatography

For small sample volumes two types of commercial desalting columns were used for buffer exchange, following the manufacture's protocol: PD-10 (up to 2.5 ml) and MiniTrap (up to 0.5 ml) G-25 columns (GE Healthcare). In the case of Tiff4631 proteins for large buffer exchange volumes, a 25 ml column was packed with the G-25 matrix. Column was calibrated obtaining a maximum volume for buffer exchange of 20-25 ml and protocol consisted on column equilibration with the desired buffer (minimum two column volumes), protein solution was loaded in up to 25 ml and recovered in about 30-35 ml in the new buffer. The process was performed in a FPLC to monitor the outcome of the protein in the new buffer (UV-detector) and the old buffer (conductivity detector).

2.4.6. Protein concentration methods

Sample volumes lower than 5 ml were concentrated by speed-vac (SVC200H, Savant), whereas samples in higher volumes (5-20 ml) were concentrated using commercial Viva-Spin 20 (GE Healthcare) concentrators following manufacture's instructions. It is important to note that commercial concentrators do not change buffer conditions (in theory), but speed-vac method concentrates proteins and also salts, thus large volumes could not be concentrated by this method as it dramatically increases the salt concentration.

2.4.7. Mass spectrometry

Instituto de Química-Física "Rocasolano" (CSIC) mass spectrometry service was used for routine check of sample homogeneity and purity after purification procedures. For protein peptide footprint, samples (cut gel bands) were sent to the mass spectrometry service of the IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire).

2.4.8. Determination of protein concentration

Protein concentration was determined by measuring sample UV absorbance at 280 nm with a NanoDrop 2000 (Thermo Scientific) spectrophotometer. Concentration was calculated using Lambert-Beer law (Equation 2.1). The theoretical molar absorbance coefficient (ε_m) was obtained from ExPASy ProtoParm web source (http://web.expasy.org/protparam/).

Eq. 2.1. $$A = \varepsilon_m \cdot l \cdot c$$

The sample concentration of the Gbp2p RRM1 construct, lacking tyrosine or tryptophan residues, was quantified by the BCA method using a commercial kit (Micro BCA Protein Assay Kit, Pierce) and following manufacture instructions. The total protein concentration in yeast crude extract was also determined with the same approach.

2.5. NUCLEIC ACID METHODS

Synthetic DNA and RNA oligonucleotides were purchased from IDT, except some preparations of large RNA molecules.

2.5.1. Oligo purification protocol

Ethanol precipitation was used to remove the excess of salt and other contaminants from synthetic DNA oligonucleotides used in PCR. Lyophilised oligonucleotides were resuspended in 180 μl of water and 20 μl of 3 M sodium acetate at pH 5.2, then 600 μl of iced ethanol were added and then, incubated at -20ºC for at least half an hour. The resulting solution was centrifuged (14000 rpms) for 10 minutes at 4ºC and the pellet was washed with room temperature 70% ethanol solution, centrifuged again (10 minutes 14000 rpms), let to dry to the air and solubilized in sterile water. All oligonucleotides used in binding studies (either synthetic DNA/RNA or larger-scale production of RNA by *in vitro* transcription) were purified utilizing commercial desalting columns (MiniTrap G-25, GE Healthcare) following the same protocol as used in protein exchange buffer.

2.5.2. Determination of nucleic acid concentration

The quantification concentration of DNA and RNA samples was determined by measuring the UV absorbance at 260 nm with a NanoDrop 2000 (Thermo Scientific) and using Lambert-Beer law (Equation 2.1). Theoretical molar absorbance coefficient (ε_m) was obtained from the commercial data sheet specifications.

2.5.3. *in vitro* RNA transcription

Some large RNAs were produced by *in vitro* RNA transcription for binding studies of Gbp2p and Hrb1p. This *in vitro* reaction was performed by homemade T7 RNA polymerase using a

pair of DNA oligonucleotides as template following the procedure described in the literature (Milligan and Uhlenbeck, 1989).

Prior to large-scale preparation (typically 1-5 ml), some test reactions (25 μl) were done to optimise the concentration of magnesium in the reaction. Transcription was performed in a buffer containing 40 mM Tris-HCl pH 8.1, 1 mM spermidine, 0.01% Triton X-100, 5 mM DTT, 5-50 mM $MgCl_2$, 4 mM each ribonucleotide triphosphate (rNTP), 50 mg/ml of T7 RNA polymerase and 1 mM of each DNA oligonucleotide template. The reaction mix was incubated for 2-3 hours at 37ºC, quenched by the addition of EDTA up to 30 mM and the resultant product was checked by urea denaturing PAGE.

Transcription product was cleaned twice by extraction with phenol / chloroform / isoamyl alcohol (250 μl per ml of reaction) and RNA was precipitated overnight with ice-cooled ethanol (2 ml per ml of reaction). After centrifugation (15 minutes at 14000 rpms), the pellet was washed with 500 μl of room temperature 70 % ethanol solution, solubilized in RNAse free water and quantified. Alternatively, a treatment with DNAse I (10 U/ml) could be performed to eliminate the DNA oligonucleotides prior to protein extraction.

2.6. IN VIVO YEAST METHODS

Several different experiments that were performed for *in vivo* functional studies of Gbp2p and Hrb1p proteins in *S. cerevisiae* are described in the following sections.

2.6.1. Growth rate assays

Two different experimental approaches were performed for comparing strain growth rate: drop tests and liquid media growth. The first test was useful to study the impact of different stresses (individually or in conjunction) on the growth rates. To perform it, each strain was grown in 1 ml of YPD at 30ºC and when they reached an OD_{600nm} of about 1, cultures were diluted to OD_{600nm} of 0.1, 0.01, 0.001 and 0.0001 with sterile water. Finally, 5 μl of each dilution were placed on plates (several different media were utilized, see section 2.1.1) and they were incubated at 4 different temperatures (thus for each assay four duplicates were done): 16, 25, 30 and 37ºC for 1-3 days.

To evaluate the growth phenotypes induced by different mutations under normal conditions and also under saline stress, assays in liquid media were done. Starting from fresh cultures, each strain was grown in YPD media until reaching an OD_{600nm} of 1, then cultures at an OD_{600nm} of 0.1 were prepared by dilution in 20 ml of appropriate media (YPD or YPD + 1M NaCl) and shook at 30ºC during at least 2 days. Cell density was monitored during this time by measuring the OD_{600nm} of each culture.

2.6.2. β-galactosidase expression tests

These assays were made in order to quantify the expression levels of β-galactosidase protein in the different Gbp2p/Hrb1p deletion mutant genetic backgrounds. This type of experiment required the transformation of the strains with plasmids that contain different versions of this reporter gen. All the plasmids (see Table 2.2) contain the uracil auxotrophy gen as genetic marker. Cultures (5 ml) of CSM-Ura media were inoculated with different strains (wild type, *gbp2pΔ*, *hrb1pΔ* and *gbp2pΔ/hrb1pΔ* mutants), previously transformed with a given plasmid, were grown at 30ºC until reaching an OD_{600nm} of around 0.5 (the exact value must be measured for every culture) and 750 μl of each one were used for β-galactosidase assay. Cells were pelleted (1 minute at 14000 rpms) and resuspended in 500 μl of buffer Z (100 mM sodium phosphate at pH 7.2, 10 mM KCl, 1 mM $MgSO_4$ and 0.36% of β-mercaptoethanol), then 200 μl of ether were added and the mix was vortex for 20 seconds. Ether was evaporated opening tubes under the hood for about 10 minutes and solution was vortex-mixed again and incubated at 30ºC for 5 minutes. Finally 100 μl of ONPG (ortho-nitrophenyl-β-galactoside) solution (4 mg/ml in buffer Z) were added to each sample and after incubation at 30ºC for 1 minute, reaction was stopped by the addition of 250 μl of a 1M Na_2CO_3 solution. Absorbance at 420 nm was measured for each sample and β-galactosidase activity (in arbitrary units) was calculated using Equation 2.2, where Vol means the original culture volume (750 μl) and Time the reaction time (1 minute) (Dreumont and Seraphin, 2013).

Eq. 2.2. $$Activity = \frac{1000 \cdot Abs_{420nm}}{OD_{600nm} \cdot Vol \cdot Time}$$

Each assay was performed using two different cultures of the same strain (two biological replicates) and two different assays for each culture (two technical replicates), calculating an arithmetic mean value for each strain with the corresponding standard deviation.

2.6.3. Reverse transcription and quantitative PCR

Quantification of the β-galactosidase mRNA was performed to complement the above protein-based activity assays. First the total mRNA was extracted from the different yeast cultures and reversibly transcribed to cDNA. The amounts of β-galactosidase and URA3 (as control house-keeping gene) cDNAs were finally determined by quantitative PCR.

Total RNA was extracted from 15 ml of yeast culture (grown as in β-galactosidase expression tests). The pelleted cells (1 minute at 14000 rpms) were resuspended in 0.5 ml of cooled RNA extraction buffer (100 mM Tris-HCl pH 7.5, 100 mM LiCl and 1 mM EDTA filtered solution) and 0.5 ml of glass beads, 10 µl of 10% SDS solution and 0.5 ml of phenol / chloroform / isoamyl alcohol mix were added. This mixture was vigorously shaken using 8 vortex cycles of 30 seconds on and 30 seconds off (on ice) to liberate the cytoplasm contents. After centrifugation (5 minutes at 14000 rpms), the aqueous phase was extracted twice more with 0.5 ml of the phenol / chloroform / isoamyl alcohol solution to further remove protein contents. The pH of the aqueous phase was adjusted to 5.0 with 50 µl of 3 M sodium acetate solution, RNA was precipitated with iced ethanol (1.5 ml) for half an hour at -80ºC, centrifuged (15 minutes at 14000 rpms) and the pellet was washed with 0.5 ml of room temperature 70% ethanol solution. The RNA pellet was resuspended in 45 ml of water and treated with 2 units of DNAse I (Ambion) in the appropriate buffer for 30 minutes at 37ºC. Finally, the RNA was purified following the same protocol used before (protein extraction with phenol / chloroform /isoamyl alcohol mix and ethanol precipitation). The purified RNA was solubilized in 100 µl of water and quantified (approximately $1OD_{260nm}$ = 40 ng/µl).

Reverse transcription was carried out using 1 µg of purified RNA per 25 µl of reaction and following kit procedure specifications (Life Technologies, Saint-Aubin). The total cDNA products could be either used straightaway in qPCR or stored at -80ºC. Finally, quantitative PCR was perfomed using SYBR Green methodology (Stratagene, Mx3000) and a LightCycler® 280 (Roche). Two pairs of previously tested primers were used, for quantification of β-galactosidase cDNA and URA3 gene cDNA. For each sample, three technical repetitions using each pair of primers were carried out. The reaction mix was prepared following the manufacture's protocol and the conditions used for PCR were a first activation step at 95ºC for 15 minutes, followed by 45 cycles of 10 seconds at 95ºC, 15 seconds at 60ºC and 20 seconds at 72ºC; quantification was made by continuous measurement of fluorescence (λ_{ex}

483 nm and λ_{em} 533 nm). The obtained crossing points or cycle threshold values (Cp) for each measurement were analysed using a relative quantification method without real-time PCR efficiency correction (Equation 2.3) obtaining a relative quantitative value (R). This value relates the β-galactosidase cDNA concentration, corrected with URA3 house keeping gene concentration, between each mutant and wild type strains (Livak and Schmittgen, 2001).

Eq. 2.3. $$R = 2^{-\left((Cp_{\beta gal} - Cp_{URA3})_{mutant} - (Cp_{\beta gal} - Cp_{URA3})_{wild\ type}\right)}$$

R values significantly lower than 1 correspond to less amount of β-galactosidase mRNA in the mutant than in the wild type strains; in contrast, higher R values than 1 mean higher β-galactosidase mRNA concentrations.

2.6.4. Tandem Affinity Purification

This experiment represents one of the breakthroughs in the field of proteomics and interactomics and consists in the purification of functional protein-protein complexes constructed *in vivo* by two consecutive affinity purification steps (Puig et al., 2001). Modified yeast strains were prepared by incorporation of TAP-tag (calmodulin binding peptide + TEV site + protein A) fused to C-terminal of target protein or mutant (see section 2.3), in order to identify the proteins that co-purify with it. Table 2.5 lists all the TAP-tag modified protein strains utilized in tandem affinity purification experiments during this work.

Strain	Description
GBP2-TAP	Gbp2p protein TAP tagged in C-terminal.
GBP2ΔRRM3-TAP	Gbp2p mutant construction without RRM3 domain TAP tagged in C-terminal.
GBP2ΔRRM12-TAP	Gbp2p mutant construction without RRM1 and RRM2 domains TAP tagged in C-terminal.
GBP2-TAP / *hrb1*Δ	Gbp2p protein TAP tagged in C-terminal, in a genetical background lacking Hrb1p.
GBP2ΔRRM3-TAP / *hrb1*Δ	Gbp2p mutant construction without RRM3 domain TAP tagged in C-terminal, in a genetical background lacking Hrb1p.
NPL3-TAP	Npl3p protein TAP-tagged in C-terminal.
NPL3-GBP2RRM3-TAP	Chimeric protein compound by Npl3p with the RRM3 domain of Gbp2p TAP-tagged in C-terminal.

Table 2.5. – List of different modified strains used for TAPs experiments.

General procedure consists on a first preparation step of protein extraction, then the first affinity purification step (binding to IgG beads), a TEV protease treatment, the second affinity purification (interaction with calmodulin beads) and a SDS-PAGE analysis followed by mass spectroscopy band identification of the co-purified proteins (Figure 2.9).

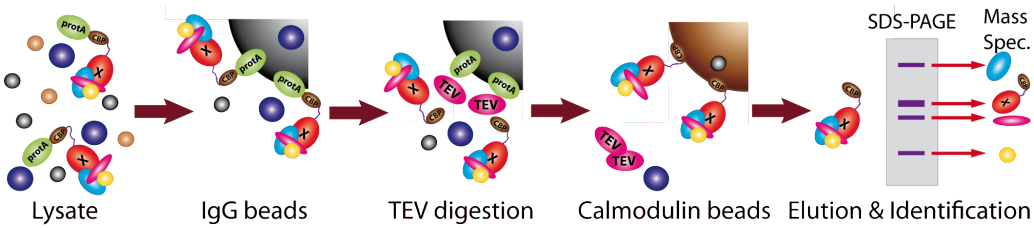


Figure 2.9. – Graphic representation of Tandem Affinity Purification procedure: lysis, IgG affinity purification, TEV protease treatment, calmodulin affinity purification and eluate analysis by SDS-PAGE and mass spectrometry protein peptide footprint.

2.6.4.1. Cell culture conditions and lysis

Usually, 4 litres of YPD were inoculated with fresh cultures of the desired strain and were grown at 30ºC, until an OD_{600nm} of 2 was reached. Then, cultures were centrifuged (4000 rpm at 4ºC for 20 minutes), washed with 500 ml of cold water and precipitated again. These pellets could be frozen in liquid nitrogen and stored at -80ºC for long-term periods (months) or directly used.

The yeast pellet was resuspended with the same volume of ice-cold lysis buffer (10 mM K-HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 2 mM benzamidine, 0.5 μg/ml leupeptine, 1.4 μg/ml pepstatin A, 2.4 μg/ml chymostatin and 17 μg/ml apropitin). This suspension was passed through the French press twice to lysate cells (1200 psi in High position) and supplemented with KCl up to a final concentration of 0.2 M. Two consecutive centrifugation runs (20500 rpm for 30 minutes and 33500 rpm for 90 minutes) at 4ºC were used to isolate a clear fraction from the lysate. This fraction was dialysed/concentrated against glycerol-containing buffer (20 mM K-HEPES pH 7.9, 50 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 25% glycerol, 0.5 mM PMSF and 2 mM benzamidine) for 3 hours at 4ºC, using a membrane with a molecular weight cut off (MWCo) of 6000-8000 Da. At this point the extract could be either frozen in liquid nitrogen and stored at -80ºC, or further processed.

2.6.4.2. IgG affinity chromatography

The extract from the previous step was supplemented with variable volumes of solution stocks of Tris-HCl pH 8.0, NaCl and detergent Igepal (Sigma-Aldrich) until reaching final concentrations of 10 mM, 100 mM and 0.1% respectively. Then, yeast extract was incubated for 2 hours at 4ºC in a column (Chromatography Columns, BIO-RAD) with 200 μl of IgG beads (IgG Sepharose G Fast Flow, GE Healthcare), previously equilibrated with IPP150 buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.1% Igepal). The flow through was recovered and beads were washed three times with 10 ml of IPP150 buffer and one final wash with 10 ml TEV cleavage buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Igepal, 0.5 mM EDTA and 1 mM DTT). Fractions were collected at each step and stored for later analysis by Western-Blotting.

2.6.4.3. TEV treatment

TAP-containing IgG beads were incubated with approximately 100 units of TEV enzyme (in 1 ml of cleavage buffer) for 2 hours at 16ºC. Beads were drained and washed with additional 200 μl of TEV cleavage buffer, both fractions, that should contain the target complex, were pooled together.

2.6.4.4. Calmodulin affinity chromatography

Three volumes of calmodulin binding buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 2 mM $CaCl_2$, 0.1% Igepal and 10 mM β-mercaptoethanol) and 3.6 μl of 1M $CaCl_2$ were added to the previous 1.2 ml fraction. This mixture was incubated for 1 hour at 4ºC in a clean column with 200 μl of calmodulin beads (Calmodulin Sepharose 4B, GE Healthcare), previously equilibrated with calmodulin binding buffer. Then, flow through was recovered and beads were washed three times with 10 ml of calmodulin binding buffer. Finally, the proteins bound to the beads were eluted with 5 fractions of 200 μl of calmodulin elution buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 2 mM EGTA, 0.1% Igepal and 10 mM β-mercaptoethanol) and one fraction of 200 μl of 1% SDS solution.

2.6.4.5. SDS-PAGE and mass spectrometry characterization

Different elution samples were dried (in SpeedVac) and then resuspended in 20 μl of loading buffer (10% glycerol, 1% SDS, 0.01% bromophenol blue and 1 mM DTT). Then, they were loaded into a 5-20% SDS-polyacrylamide gradient gel (as described in PAGE section). Gel bands requiring identification by peptide footprint were cut and sent to mass spectrometry service. If the gel was silver stained, a preparative destaining step was performed before sending the products: cut bands were incubated in a 0.5 ml of 0.2% potassium ferricyanide ($K_3[Fe(CN)_6]$) and 0.02% sodium thiosulphate ($Na_2S_2O_3$) solution for 15 minutes and then gently washed with milliQ water until no background colour was observed.

2.6.5. Preparation of yeast extract for *in vitro* cross-linking

The last functional experiment carried out with yeasts consisted on a cross-linking reaction between purified recombinant proteins and yeast extract proteins. This section describes the protocol to obtain yeast lysates for this purpose. The cross-linking and protein expression and purification sections of this chapter describe additional details for the *in vitro* cross-linking analysis.

Cell pellet coming from 200 ml culture (obtained as in previous protocols) were resuspended in 1 ml of lysis buffer (10 mM K-HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM PMSF, 2 mM benzamidine, 0.5 μg/ml leupeptin, 1.4 μg/ml pepstatin A, 2.4 μg/ml chymostatin and 17 μg/ml apropitin) and 1 ml of glass beads were added. Lysis was carried out mechanically using vortex (10 times 15 seconds on, 45 seconds off) and lysates were clarified by the usual two consecutive centrifugations (8000 rpm for 10 minutes and 53000 rmp for 30 minutes). At this point and after quantified protein concentration the extracts could be stored at -80ºC for further crosslinking experiments.

2.7. POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE)

Polyacrylamide gel electrophoresis was routinely used along this work for different purposes:

- Denaturing PAGE that separates the biomolecules according to their size. Normally, 15% SDS-PAGE mini-gels were used to monitor different purification steps and to verify protein homogeneity and purity. For large proteins SDS-PAGE gels with less percentage of

acrylamide (10%) or gradient SDS-PAGE gels were used. Urea denaturing nucleic acid PAGE was used during RNA preparation and for nucleic acid quality control.

- Native PAGE that separates biomolecules depending on their size, charge and oligomerization state. The composition of native protein gels is similar to the above but without SDS. The same gels were used on nucleic acid applications and in particular in telomeric DNA analysis.

Protein PAGE gels were routinely stained with coomassie and with silver for detection of proteins at low levels (ng-pg). For even higher sensitivity inmunodetection (western blotting) was performed. Nucleic acid development was performed by using toluidine staining. Finally, DNA/RNA-protein interactions were studied by electrophoretic mobility shift assays (EMSA) or their version using fluorescent probes (f-EMSA) that has higher sensibility.

2.7.1. Protein PAGE

For routinely analysis, samples (containing loading buffer: 10% glycerol, 1% SDS, 0.01% bromophenol blue and 1 mM DTT or 1% β-mercaptoethanol) were heated at 90ºC for 5 minutes and loaded into homemade classic SDS-polyacrylamide biphasic gels. The separating gel occupied about ¾ at the lower part and were prepared at 15%, 12% or 10% of polyacrylamide, 0.45 M Tris-HCl pH 8.8, 1% of SDS, 1% of ammonium persulphate and 0.02% of N,N,N,N-tetramethylethylendiamine (TEMED). The stacking gel is cast in the ¼ upper portion and consisted of 5% of polyacrylamide, 0.125 M Tris-HCl pH 6.8, 1% of SDS, 1% of ammonium persulphate and 0.02% of TEMED. In some cases, commercial polyacrylamide gels were used: 12% of polyacrylamide gels and gradient gels from 4-20% and 4-15% of polyacrylamide (Mini-PROTEAN TGX, BIO-RAD). Electrophoresis was performed immersing the cathode and the anode in Laemmli buffer (25 mM Tris Base, 200 mM glycine and 1% SDS) and connecting the electrophoresis cell to a power source at about 25 mA (150-250 V) for 30-45 minutes or until the dye run out of the gel.

Two different protein ladders were used for molecular weight comparison, a homemade standards mix (phosphorylase b 97.3 kDa, bovine albumin 66 kDa, chicken egg albumin 44 kDa, carbonic anhydrase 18 kDa, trypsin inhibitor 18 kDa, lysozyme 14.6 kDa and ubiquitin 8.6 kDa) and a commercial one (Precision Plus Protein Standars, BIO-RAD) that includes pre-stained proteins.

For native protein PAGE, the samples (in loading buffer without SDS) were loaded into native polyacrylamide gels (without heating step), these gels were prepared with 15% of polyacrylamide, 0.45 M Tris-HCl pH 8.8, 1% of ammonium persulphate and 0.02% of TEMED. Electrophoresis cell was performed in the cold room (4ºC) with pre-chilled running buffer (25 mM Tris Base and 200 mM glycine) and at low voltage of 100 V for one hour. All proteins analysed by native PAGE moved to the positive electrode (the same mobility as SDS-protein).

2.7.1.1 Coomassie blue staining

This method was routinely performed following this protocol:

- Step 1: Incubate the gel with 0.1% commassie R250 brilliant blue (Sigma) solution (50% ethanol and 10% acetic acid).
- Step 2: Destain the gel with a 5% ethanol and 7.5% acetic acid buffer.

2.7.1.2. Silver staining

This method was used in applications that require higher sensitivity and follows the protocol:

- Step 1. Incubate the gel for at least one hour in a fixing solution (50% methanol, 12% acetic acid and 0.02% paraformaldehyde).
- Step 2. Wash the gel three times with 50% methanol for 8 minutes each.
- Step 3. Pretreat the gel with a 0.02% sodium thiosulphate solution for 1 minute.
- Step 4. Wash the gel with water and incubate it with a 0.2% silver nitrate solution for 20 minutes.
- Step 5. Wash the gel again with water and develop it using a fresh 6% sodium carbonate, 0.02% sodium thiosulphate and 0.02% paraformaldehyde solution until achieving the desired level of staining.
- Step 6. Stop development by immersing the gel into the stop solution (50% methanol and 12% acetic acid).

2.7.2. Western-blotting

This technique was used for selective detection of one protein, normally TAP-tagged proteins in yeasts extracts. The protocol comprises the following steps:

- Step 1. A standard SDS-PAGE gel was performed, usually an 8% gel.

- Step 2. The gel was framed on a nitrocellulose membrane (Amersham Biosciences) using a horizontal electrophoresis cell. The gel was carefully put over the transfer membrane and both sandwiched between several pieces of filter paper and two scouring pads. Assemble of this transfer cassette was performed with the components immersed in cooled running buffer (25 mM Tris Base, 200 mM glycine and 20% ethanol). Then, the cassette was placed in the electrophoresis cell with the gel facing the cathode and the membrane the anode.
- Step 3. Electrophoresis was run at 4ºC (in the cold room) using a voltage of 80 V for 1-2 hours.
- Step 4. After disassembling the gel-membrane cassette, the membrane was first stained with a solution of Ponceau red (2% in a 1% acetic acid aqueous solution) for global protein staining (to check if transference was right).
- Step 5. The membrane was blocked with about 20 ml of a milk solution (5% non-fat dried milk in 100 ml of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4) with 0.05% of Tween 100X) for 1 hour at room temperature.
- Step 6. A small volume (10 ml) of milk solution with PAP antibody (Sigma) at 1:3000 dilution was added for another incubation of at least one hour.
- Step 7. The membrane was washed three times using PBS with 0.05% Tween 100X for 10 minutes each.
- Step 8. It was developed using a peroxide commercial kit (Immun-Star WesternC kit, BIO-RAD) in an imaging system (Molecular Imager Chemi DOC XRS+, BIO-RAD).

2.7.3. Nucleic acids PAGE

Two different types of polyacrylamide gels were used for nucleic acids one in native conditions and other in denaturing ones. The denaturing PAGE was used during RNA transcription and purification and the native PAGE was performed for different nucleic acids analysis.

Urea denaturing gels were prepared typically at 20% of polyacrylamide on TBE buffer (5.5 g/l boric acid, 10.8 g/l Tris-Base and 0.3 g/l EDTA) with 8 M urea, 1% of ammonium persulphate and 0.02% of TEMED for polymerization. Native PAGE gels were prepared at 15 or 20% of polyacrylamide on TB buffer (5.5 g/l boric acid, 10.8 g/l Tris-Base) with 1% of ammonium persulphate and 0.02% of TEMED.

As in the protein case, denaturing gels were run at room temperature, using TBE buffer as running buffer, loading the samples with urea (loading buffer: 0.4 mg/ml of xylene cyanol (Sigma-Aldrich), 0.4 mg/ml bromophenol blue (Sigma-Aldrich), 10% glycerol and 6 M urea) and performing the electrophoresis at 200-300 V for about 30 minutes. Native PAGE was performed in the cold room (4ºC) using TB buffer, mixing the samples with the same loading buffer described before but without urea and running the electrophoresis with a lower voltage (about 100-150V for 1-2 hours).

2.7.4. EMSA and f-EMSA

Electromobility gel Shift Assays (EMSA) were used in the analysis of protein-DNA interactions. This technique identifies different species (monomeric, oligomeric, protein-bound state, etc.) of a given oligonucleotide probe on the bases of the differences of their mobility in native PAGE. The method is referred as f-EMSA when fluorescence-labelled oligonucleotides (6-carboxyfluorescein linked to the 3' phosphate) were used. The native gels were prepared at 10% of polyacrylamide in 0.45 M Tris-HCl pH 8.8, 1% of ammonium persulphate and 0.02% of TEMED. The buffer and run protocol were similar to those described for other native gels.

Regular oligonucleotides bands in the gels were detected by incubating with a warm toluidine blue (0.1% w/v in water) solution, and later destained with water. This dye binds preferentially to nucleic acids, although in a few cases, some level of protein staining was observed. To complement this information the gel was later stained with Coomassie (see protocol above), which theoretically binds to the protein bands. More specific information was obtained using 6-carboxyfluorescein labelled DNAs. The f-EMSA does not substantially differ from the standard EMSA but nucleic acids band detection was made using the imaging system (Molecular Imager Chemi DOC XRS+, BIO-RAD). The gel was trans-illuminated with an UV excitation source of λ=302 nm and fluorescence emission was detected by using a standard white light filter.

2.8. ANALYTICAL GEL FILTRATION

Analytical gel filtration chromatography was performed using a Superdex 200 column (10/300 GE Healthcare), mounted on an FPLC system (ÄTKA, Amersham Biosciences) for chromatogram monitoring. The mobile phase used for all experiments (including standards)

was composed by 25 mM potassium phosphate (pH 6,5), 25 mM NaCl and 1 mM DTT. Different constructs of Hrb1p and Gbp2p proteins were assayed by following an identical procedure: 100 μl of protein solution at 100 μM protein [in 25 mM potassium phosphate (pH 6,5), 25 mM NaCl and 1 mM DTT buffer)] were loaded in the column and then a flow of 0.5 ml/minute was applied during 30 ml. Absorbance at 280 nm was monitored to detect the protein peaks.

The results were compared with a previously constructed calibration dataset with the following standards: aprotinin (6,7 kDa), ribonuclease A (13,7 kDa), myoglobin (16,7 kDa), deoxiribonuclease I (29 kDa), albumin (44 kDa) and bovin serum albumin (BSA 67 kDa).

2.9. CIRCULAR DICHROISM SPECTROSCOPY

Circular dichroism spectroscopy experiments were recorded in a JASCO J-810 spectropolarimeter equipped with a Peltier system for temperature control and the results were registered and analysed using Spectra Manager software. Normally, each spectrum was obtained by adding three scans; the data was acquired at a rate of 20 nm/minute with a time response of 4 seconds, recording ellipticity values each 0.2 nm.

In protein experiments, about 200 μl of each sample were prepared at a concentration of 0.2 mg/ml in a 25 mM potassium phosphate (pH 6,5), 25 mM NaCl and 0,1 mM DTT buffer and all measurements were performed using a 1 mm path length quartz cuvette. For temperature stability measurements, an exploratory set of far-UV protein spectra were acquired at 20, 35, 50, 70 and 90 ºC, from 250 to 200 nm. Then, the actual protein denaturalization curves were registered at constant wavelengths (203 and 217 nm) from 20ºC to 90ºC at a rate of 60ºC/hour. All the spectra were corrected by subtracting equivalent data obtained from blank samples composed of buffer. The total ellipticity values were transformed into molar ellipticity per residue, following the Equation 2.4, where θ is ellipticity, $[\theta]_{MR}$ means molar ellipticity per residue, C_{MR} is molar concentration per residue and l is the path length cuvette in meters (0.001 m):

Eq. 2.4. $$\theta = [\theta]_{MR} \cdot C_{MR} \cdot l \qquad C_{MR} = \frac{C_{prot}(mg/ml)}{MW} \cdot n^{o}_{aa}$$

The CD spectra of DNA were acquired from 320 to 220 nm at 25ºC using the same 1 mm path length cuvette. Two different DNAs were studied: TG-43 and TG-36; in different buffers: 20 mM Tris-HCl (pH 8.0) and from 0 to 150 mM NaCl or KCl.

Protein-DNA interaction studies were made by recording the CD spectra of mixtures between TG-43 and TG-36 DNAs, and Gbp2p RRM12 protein at DNA:protein ratios of 1:0, 1:1, 1:2, 1:4 and 1:5. Salts and proteins were added to DNAs in different order, obtaining different results. Spectra were corrected subtracting the CD spectra of the corresponding buffers without DNA and protein components.

2.10. FLUORESCENCE SPECTROSCOPY

2.10.1. Tryptophan fluorescence spectroscopy

Intrinsic protein fluorescence was measured at 20ºC using a FluoroMx-4 spectrofluorometer (Horiba, Jobin Ybon), controlled by FluorEssence 2.1 software. Samples were prepared in a 3 x 3 mm path length quartz cuvette using a 25 mM potassium phosphate at pH 6.5, 25 mM NaCl and 0.01 mM DTT buffer and protein concentrations were set to be equivalent to an absorbance of 0.03 uA at 280 nm. The excitation wavelength was 290 nm and the emission spectra were recorded form 450 to 300 nm. Slit widths for excitation and emission were 3 and 2 nm, respectively.

2.10.2. Fluorescence anisotropy measurements

The analysis of protein-DNA and protein-RNA interaction by fluorescence anisotropy measurements was performed at 20ºC on a PC1 photon counting steady-state ISS spectrofluorometer (Champaign, IL, USA). The excitation (λ_{ex}) and emission (λ_{em}) wavelengths were 495 nm and 520 nm respectively, the slit widths for excitation and emission were fixed to 1 and 3.16 nm respectively and experiments were carried out in a 3 x 3 mm path length quartz cuvette. Each of the seven to eight titration points in the binding isotherms correspond to the mean of 3 independent experiments. Once in the spectrofluorimeter each sample was repeatedly measured and the anisotropy values reported represent the average of the last 8-10 values obtained after reaching thermal equilibrium.

Gbp2p and Hrb1p experiments were performed in a 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.01 mM DTT buffer with concentrations of 20 nM of 6-carboxyfluorescein labelled TG-43 or TG36 DNAs. Titrations were done with Gbp2p and Hrb1p RRM12 and RRM13 constructs (Table 3.8) at DNA:protein ratios from 1:0 to 1:1000.

Protein-RNA interaction studies of wild type Nab2p constructs and mutants (Table 4.2 and 4.3) were performed with the same hardware settings than above. The concentration of 3'-end fluorescein-labelled RNAs in the titrations was always 22 nM and the buffer used consisted of 25 mM Tris HCl pH 8.1, 100 mM NaCl, 0.1 mM DTT. Competition experiments were conducted under equivalent conditions using 20 μM of competitor nucleic acid.

Analysis of the isotherms of binding was performed using BIOEQS (Rosales and Royer, 2008; Royer et al., 1990). This software allows recovering the free energies of formation of the postulated complexes from their individual elements by means of a numerical solver engine. The errors in the free energies retrieved were calculated by rigorous confidence limit testing at the 67% level. A model based on a single (1:1) RNA/protein complex was fit to the isotherms recovered from titrations of fluorescently labelled RNA with protein. For the competition titrations, two (1:1) complexes were postulated, consisting of protein and labelled or unlabelled RNA, respectively. In the analysis, the free energy of formation of the protein/labelled RNA complex and its anisotropy were fixed to the values obtained in the absence of competitor RNA.

2.11. *In vitro* CROSS-LINKING EXPERIMENTS

Protein-protein interactions were also analysed by *in vitro* cross-linking experiments. The general procedure consists on mixing two protein interacting candidates in the presence of a reactive chemical agent capable to react with protein chemical groups and the analysis of adducts appearance by SDS-PAGE analysis and eventually by mass spectrometry. Two different cross-linker reactants were used in this work: glutaraldehyde (Sigma) and sulfo-NHS-LC-diazirine (Pierce) (Figure 2.10). The different protocols used for each cross-linking reaction are described in the following sections.

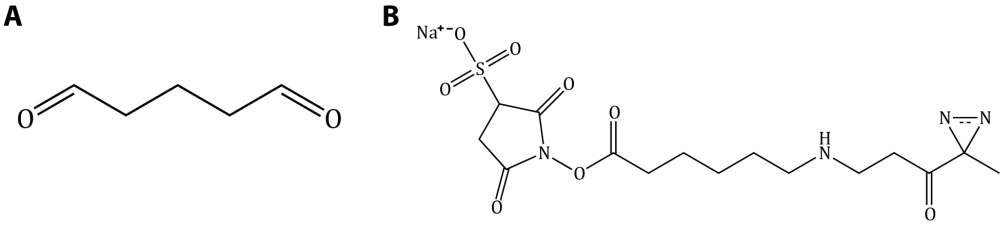


Figure 2.10. – Chemical representation of both cross-linkers: A) glutaraldehyde (1,5-pentanedial) and B) sulfo-NHS-LC-diazirine, (sulfosuccinimidyl 6-(4,4'-azipentanamido)hexanoate).

2.11.1. Glutaraldehyde cross-linking

Glutaraldehyde (GLH) is a homobifuntional, highly reactive agent widely used in protein cross-linking reactions and in protein immobilization. This chemical reacts with different nucleophile groups present in proteins, generally with the ε amino group in lysine residues. (Migneault et al., 2004). To minimize the non-specific background generated due to the high reactivity of GLH, a methodology based on cross-linker vapour diffusion to the protein mix was performed (Fadouloglou et al., 2008). For each crosslinking experiment, 200 μl of different Tif4631p/Pub1p mixtures were prepared in 25 mM potassium phosphate pH 8.0, 150 mM NaCl and 0.1 mM DTT buffer in 0.5 ml tubes. Protein concentrations were normalized according to their number of lysines to compensate the larger number of these reactive groups in Tif4631p protein. Pub1p constructs (RRM12 and RRM3) were prepared at 10 μM, Tif4631p (1-402) at 1 μM and Tif4631p (1-184) at 4 μM. Next, a 5 μl drop of 0.5 M aqueous GLH solution was placed on the lid of the tube, allowing the GLH vapours to diffuse slowly to the solution. The tubes were incubated (closed) at 4ºC for 16 hours and the reaction was quenched by adding 10 μl of 2 M Tris-HCl pH 8.0 solution. Adduct formation was detected as a thin band, in the SDS-PAGE, larger than the two reactive proteins and confirmed by mass spectrometry. Control reactions with the two proteins separated and without GLH were performed simultaneously to the main reaction and under the same conditions.

2.11.2. Hetero-bifunctional cross-linking

The use of sulfo-NHS-LC-diazirine (Pierce) (Figure 2.10) as cross-linker agent results in a cleaner, less sophisticated and more specific reaction than the above method. This compound contains an N-hydroxysuccinimide ester (NHS) group in one end and a diazirine ring in the other. The NHS part reacts with primary amines (-NH_2), present in amino acid side chain

groups (*e.g.* lysine) and the diazirine ring could be activated with long-wave UV light (365 nm) creating highly reactive carbine intermediates that easily form covalent bonds with any amino acid side chain or peptide backbone. The experimental approach involves a first reaction between one protein and the cross-linker (through the NHS group) and then, after quenching the non-reacted cross-linking agent, a second reaction between the activated protein and the partner protein (or protein extract) induced by irradiation with UV light (Figure 2.11).

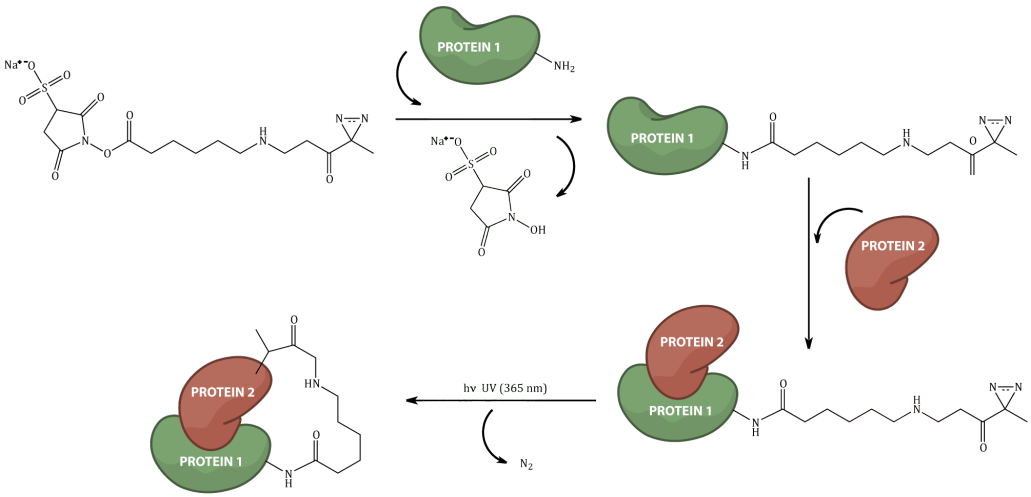


Figure 2.11. – Schematic representation of the designed heterofunctional cross-linking experiment.

Typically, first reaction was carried out preparing 100 µl of a 50 µM protein solution (Pub1p, Gbp2p or Hrb1p constructs) in PBS with 500 µM of cross linking agent. This reaction was incubated at room temperature for 30 minutes and then quenched by the addition of 2 µl of 2 M Tris-HCl pH 8.0 solution. The second reaction was performed either using purified recombinant proteins (Tif4631p constructs) or yeast extracts obtained from TAP-modified strains. In the first case, activated protein with the cross-linker was incubated for one hour with the partner protein, both at the same concentration (normally 25 µM each) under UV light (365 nm) at 4ºC. The reactions were analysed by SDS-PAGE and by mass spectrometry. For yeast extract cross-linking, 50 µl of clear lysate (at a protein concentration of about 15 mg/ml) were incubated with different volumes of activated protein (1-10 µl of 50 µM protein solution) for 1 hour, under UV light and at 4ºC. In this case, results were analysed by western.

2.12. NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY

Nuclear Magnetic Resonance Spectroscopy was used in this work with two different purposes: to obtain high-resolution structural information (either 3D structures of protein domains or secondary structure determination) and to study biomolecular interactions (protein-nucleic acids or protein-protein).

Three-dimensional structures were calculated for Gbp2p RRM3 domain and for Hrb1p RRM1, RRM2 and RRM3 ones. And many titration experiments were performed involving different constructs of all studied proteins in this work. The different protein samples were normally prepared in 25 mM potassium phosphate (pH 6.5-7.0), 25 mM NaCl and 0.1 mM DTT buffer at protein concentrations from 50-1000 μM. All samples contained 10% or 100% D_2O and 10 μM of sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) as an internal proton chemical shift reference. Heteronuclear experiments were referenced by multiplying this referenced proton frequency with the corresponding $^{1}H/X$ ratios (0.10132912 for ^{15}N and 0.2514495 for ^{13}C). NMR experiments were recorded on Bruker AV600 and AV800 MHz spectrometers equipped with cryoprobes and operated with TopSpin 2.1 software (Bruker). NMR data were processed with NMRPipe program (Delaglio et al., 1995) and viewed and analysed with CcpNMR Analysis (Vranken et al., 2005).

The next sections describe the NMR experiments used along this work as well as the methodology followed for ^{1}H, ^{13}C and ^{15}N nuclei assignment and protein structure calculation.

2.12.1. NMR experiments

NMR experiments employed in this work can be classified in several groups:

- ^{1}H monodimensional and various 2D Heteronuclear Simple Quantum Correlation (HSQC) spectra were the basic experiments to check sample integrity, to compare different constructs of the same protein and to study interactions by titration experiments.
- 3D triple (and double) resonance experiments were recorded for assignment of backbone and for side chain resonances.
- 2D Nuclear Overhauser Effect Spectroscopy (NOESY) experiments were used to obtain the distance constrains required for tridimensional structure calculation.

2.12.1.1. 1D ^{1}H and 2D ^{1}H-X HSQC experiments

Proton 1D experiment is the simplest NMR spectra and shows all observable proton signals. It is highly overlapped in proteins and therefore of limited use, but nevertheless provides a quick diagnosis about sample purity, aggregation state, degree of structure (or lack of it). Furthermore, for isolated signals in the edges (*e.g.* methyl groups or tryptophan indols) the proton 1D is a very sensitive approach to monitor interactions.

The 2D Heteronuclear Simple Quantum Correlation (HSQC) experiments have been widely used and are often considered as protein fingerprint spectra. These are based on two INEPT type magnetization transfer steps between proton and directly attached heteroatoms (^{13}C or ^{15}N) via large scalar coupling constant ($^{1}J_{H\text{-}C}$ =130-220 Hz or $^{1}J_{H\text{-}N}$ = 90-100 Hz) separated by a t1 chemical shift evolution periods (Bodenhausen and Ruben, 1980). Both types of experiments normally require protein isotope labelling with ^{15}N and ^{13}C respectively.

The majority of correlations observed in the ^{1}H-^{15}N HSQC experiment correspond to backbone amide groups (all residues except prolines) and side chains groups of asparagine (Nδ2), glutamine (Nε2), tryptophan (Nε1) and arginine (Nε). Other possible correlations involving arginine (Nη11, Nη12, Nη21 and Nη22), lysine (Nζ) and histidine (Nδ1 and Nε2) are rarely observed as their attached protons exchange very fast with the solvent. Their eventual observation is always correlated to the existence of special structural features.

The ^{1}H-^{13}C HSQC experiment in proteins contains backbone ^{1}Hα-^{13}Cα and side chain ^{1}H-^{13}C correlations. This is a more complex spectrum and displays each type of side chain (*e.g.* prolines, methyl groups, threonine Cβ-Hβ, aromatic, etc.) on specific regions. This characteristic information was useful in side chain assignment.

2.12.1.2. Triple resonance for protein backbone assignment

Protein backbone assignment was made by combining the information coming from six different types of experiments: 3D HNCA, 3D HN(CO)CA, 3D HNCO (Grzesiek and Bax, 1992c), 3D HN(CA)CO (Clubb et al., 1992), 3D CBCANH (Grzesiek and Bax, 1992b) and 3D CBCA(CO)NH (Grzesiek and Bax, 1992a). These experiments require double isotopic labelling (^{15}N and ^{13}C) of the protein. Magnetization transfer pathways (Figure 2.12) start on the proton amide for the out-and-back experiments (3D HNCA, HNCO, HN(CA)CO and HN(CO)CA), it is transferred to ^{15}N amide via a INEPT and then to a ^{13}C (α or carbonyl depending on the

carrier position) to record its chemical shift (t_1). Incorporation of an additional relay step prior t_1 evolution renders the HN(CA)CO/HN(CO)CA pulse sequences. After recording the ^{13}C chemical shift, the magnetization is driven back to the ^{15}N amide by reverse INEPT; its chemical shift is recorded (t_2) and finally a second reverse INEPT transfer it to the proton amides for acquisition (t_3). In the out experiments (CBCA(CO)NH and CBCANH) the pathway started on the side-chain protons and was transferred to ^{13}C via INEPT module. The carbon chemical shifts were encoded during a constant time evolution period (t_1), then it was transferred to amide ^{15}N for a second chemical shift evolution period (t_2) and later to the amide proton for detection (t_3).

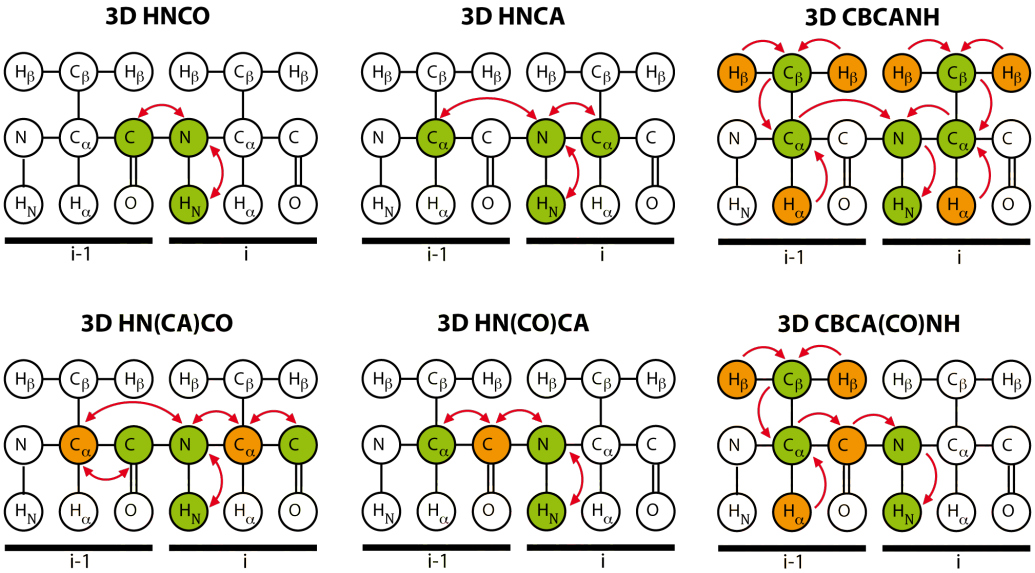


Figure 2.12 – Magnetic transferences in the six triple resonance experiments correlating nuclei of two sequential amino acids. The nuclei whose chemical shifts are recorded in the different spectra are shown in green and those nuclei that only take part in magnetic transfer pathway appear in orange. Arrows indicate the direction of the transfers (single headed for out experiments and double headed for out-and-back versions).

2.12.1.3. Side chain assignment

The experiments acquired for side chain nuclei assignment were the ^{1}H-^{13}C HSQC, some two-dimensional (TOCSY) and three-dimensional experiments (HCCH-TOCSY) for aliphatic chain assignment and two more experiments used in aromatic ring assignment: 2D CB(CGCD)HD and 2D CB(CGCDCE)HE.

2D TOCSY (Total Correlation Spectroscopy) is an proton-proton correlation experiment (Brunschweiler and Ernst, 1983) based on magnetization transfer between protons via scalar coupling constants ($^3J_{H\text{-}H}$) by applying a spin lock field (50-60 ms). In theory, the experiment correlates all proton nuclei in the same spin system (Figure 2.13.A). This experiment does not require isotopic labelling and could be acquired in H_2O and in D_2O (observing or not the water exchangeable protons). However the efficiency of the transference is strongly affected by the molecular correlation time (τ_c) and in practice this experiment is only useful for proteins with MW < 15-20 KDa.

The 3D HCCH-TOCSY experiment (Olejniczak et al., 1992) is related to the above 2D TOCSY but with some advantages. First, it shows a reduced-overlap because of the third dimension; in addition the first chemical evolution period can record either protons or carbons, generating two versions of the experiment 3D HC(C)H-TOCSY and 3D (H)CCH-TOCSY having $^1H\text{-}^{13}C\text{-}^1H$ or $^{13}C\text{-}^{13}C\text{-}^1H$ dimensions respectively (Figure 2.13.B). Secondly, the mixing is applied on the carbon (12 ms DIPSY–type supercycle for optimal transfer along the entire side system), which is more efficient than proton-proton spin-lock. Finally the $^1H\text{-}^{13}C$ transfers are achieved by short and efficient forward/backwards INEPT modules.

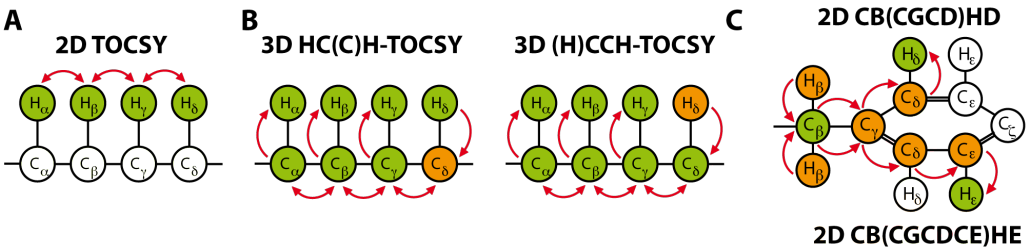


Figure 2.13. – Magnetic transferences in the 2D (A) and 3D (B) TOCSY experiments and in the aromatic ring assignment experiments (C). As in Figure 2.12, nuclei are drawn in green if they are observed in spectra and orange nuclei are those involved in magnetic transference but not registered on spectra.

Finally, 2D CB(CGCD)HD and 2D CB(CGCDCE)HE experiments (Yamazaki et al., 1993) were used to connect aliphatic and aromatic ring spin system for Phe, Tyr, Trp and His. In these experiments, magnetization is transferred from Hβ to Hδ or Hε in the aromatic ring through several relay steps whose durations are set according the corresponding scalar coupling constants (Figure 2.13.C). In the spectra a cross peak is observed at chemical shifts of Cβ and corresponding Hδ or Hε (Tyr and Phe) of the same aromatic residue. These experiments and the tridimensional HCCH-TOCSY ones require ^{13}C isotopic labelling.

2.12.1.4. NOESY

The two-dimensional ^{1}H-^{1}H NOESY experiment (Kumar et al., 1980) have off-diagonal crosspeaks correlating protons in close proximity (~ <6.0 Å). These crosspeaks are built up by a dipole-dipole transfer of magnetization through space (Nuclear Overhauser Effect) rather than through chemical bonds (as all previously experiments). The NOE build-up curve is strongly biased on type of mechanism dominating the transfer. At low mixing times, the effect grows linearly with the mixing time and the crosspeak intensities are proportional to the distance ($1/r^{6}$) between the neighbouring protons (or groups of degenerated protons like methyl) in the structure, therefore the information extracted from this experiment can be transformed into distance constrains. However, at long mixing times the spin diffusion mechanism dominates, and the intensities has to be interpreted with caution. Spin diffusion occurs by a relayed transfer of magnetization through usually dense protons network in biomolecules (nucleic acids and deuterated protein are the exception) and represent an increasingly efficient transfer mechanism for large molecules. 50 to 80 ms mixing times were employed to minimise this effect. In addition other mechanism such as chemical exchange can also give rise to crosspeaks in the spectra. The NOESY experiment, as in the case of 2D ^{1}H-^{1}H TOCSY, does not required $^{15}N/^{13}C$ isotope labelling and was acquired for protein samples in H_2O and in D_2O to distinguish water exchangeable protons.

2.12.2. Protein assignment

The protocol followed for protein assignment involved a first step of backbone nuclei assignment using triple-resonance spectra (Figure 2.14) and then the side chain assignment was completed using mainly 3D HCCH-TOCSY data sets (Figure 2.15).

The assignment process was done semi-automatically using different tools implemented in CcpNMR Analysis software suite. Typically, assignment begins with the identification and classification of spin systems in the triple resonance experiments. Each NH group observable in ^{1}H-^{15}N HSQC is associated with intraresidue (i) Cα, Cβ and CO and sequential (i-1) Cα, Cβ and CO chemical shifts. Then, these values are matched from one NH group to the next (or previous) constructing a continuous path (Figure 2.14). Some Cα and Cβ chemical shifts identify residue types unambiguously (*e.g.* Gly, Ser, Thr and Ala), thus when several of these types are found within the above walk path, an specific assignment (residue number in the protein chain) can be established.

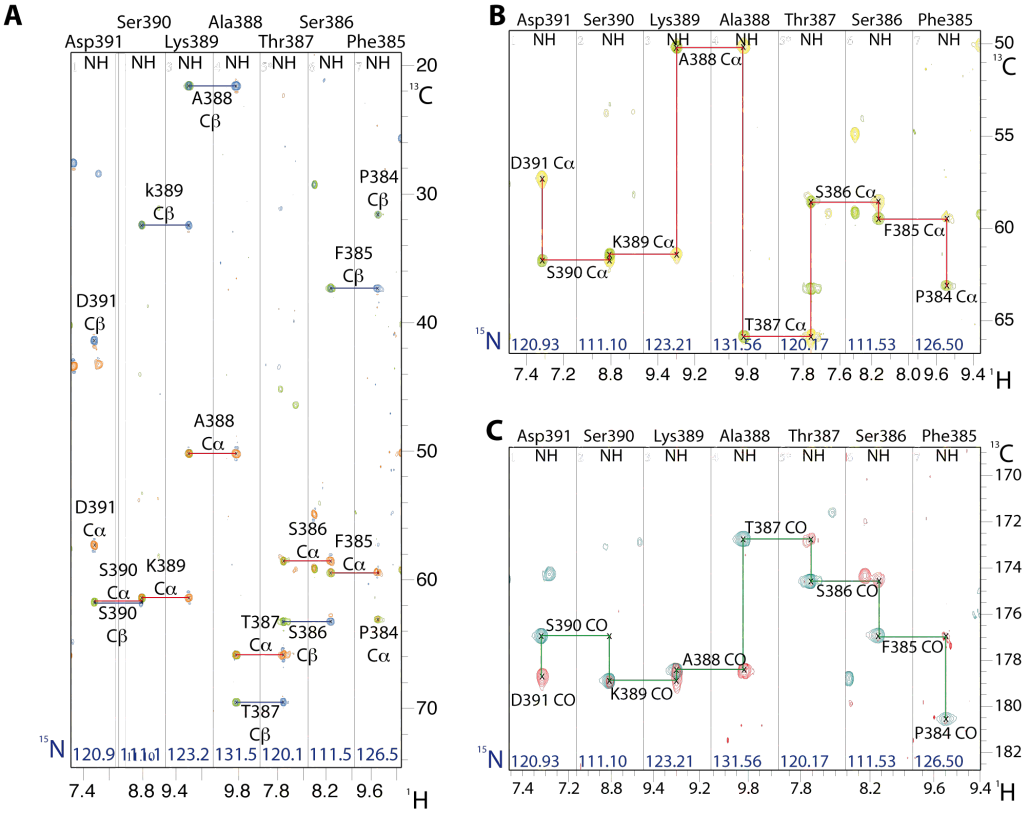


Figure 2.14. – Example of sequential backbone assignment of a seven-amino-acids segment of Hrb1_RRM3 protein. Each strip in the spectra corresponds to the ^{15}N plane for a spin system in different ^{13}C regions. A) Superposition of 3D CBCANH and 3D CBCA(CO)NH spectra. Sequential connectivities are shown by lines (orange lines for Cα and blue ones for Cβ). The two green peaks observed on each strip of the 3D CBCA(CO)NH spectrum correspond to sequential (i-1) connections to Cα and Cβ nuclei. Intra and sequential Cα (orange) and Cβ (blue) peaks could be found in 3D CBCANH spectrum. B) Overlay of the Cα region of 3D CBCA(CO)NH (green) and 3D HNCA (yellow) showing the assignment walk path with red lines. C) An equivalent comparison with the 3D HNCO (blue) and 3D HN(CA)CO (red) spectra. Connections are shown with green lines.

In the next step, the Cα and Cβ nuclei are linked with the rest of side chain nuclei by using ^{1}H-^{13}C HSQC, ^{1}H-^{1}H TOCSY and specially 3D HC(C)H and (H)CCH-TOCSY experiments. Peaks in the ^{1}H-^{13}C HSQC spectrum correlate ^{13}C and ^{1}H nuclei directly bound and the ^{1}H-^{1}H TOCSY does the same for all protons within the same spin system. Since these experiments usually have strong signal overlap, three-dimensional TOCSY were acquired to confirm/complete the assignment (Figure 2.15).

Finally, assignment is completed using ^{1}H-^{1}H TOCSY and 2D CB(CGCD)HD and 2D CB(CGCDCE)HE experiments that connect Cβ with ring protons Hδ and Hε in aromatic amino acids. Additionally, NOESY experiments (in H_2O and D_2O) were of help to assign some difficult nuclei.

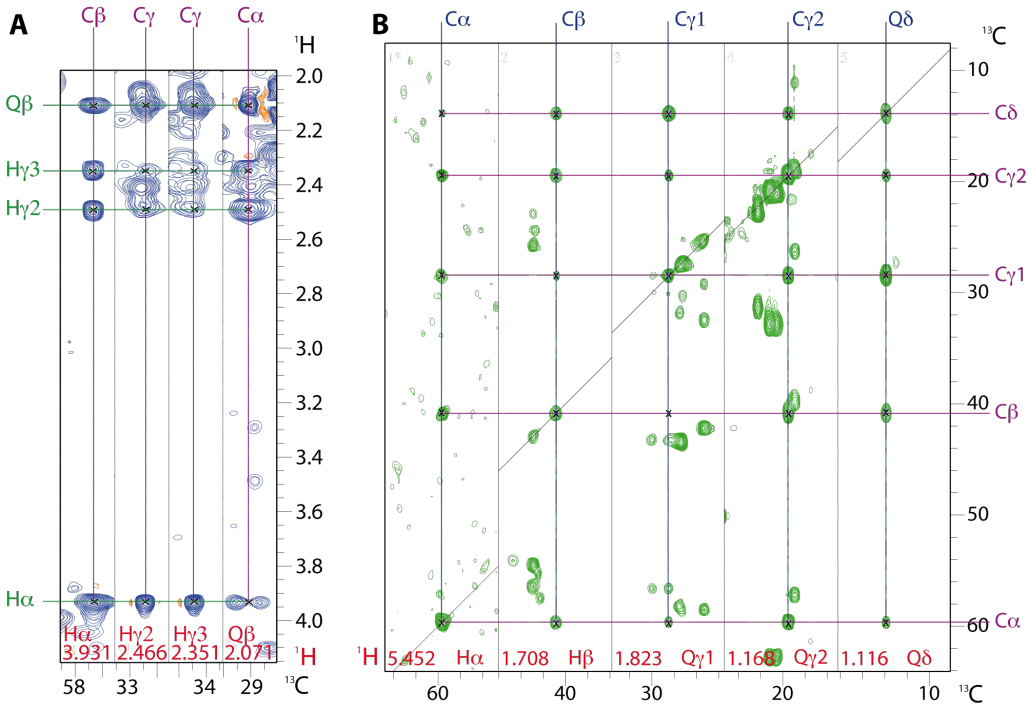


Figure 2.15. – A) Strips of 3D HC(C)H-TOCSY experiment showing all crosspeaks of Hrb1p RRM3 407 glutamine side chain spin system. B) 3D (H)CCH-TOCSY spectra strips of Gbp2p RRM3 448 isoleucine side chain spin system.

2.12.3. Structure calculation

The three-dimensional structures of various constructs of Hrb1p and Gbp2p proteins were obtained from NMR-derived angular and distance restraints.

Backbone angle restraints were obtained from the statistical comparison of the Cα, Cβ, CO, N_H and H_N chemical shift with those in an internal database of the program TALOS+ (Cornilescu et al., 1999) resulting in a prediction of range of values for ϕ and ψ backbone angles for each residue. Distance restraint lists were obtained from NOESY spectra: assigned peak lists from H_2O and D_2O two-dimensional NOESY experiments were analysed by CYANA 2.1 program

(Guntert et al., 1997) to transform peak intensities into nuclei distances, then these data were depurated in order to eliminate redundant and erroneous constraints. Both distance and angle constraint lists were implemented into CYANA 2.1 to calculate 100 random structures and the 20 lowest target function conformers were subjected to restrained molecular dynamics simulation with the program AMBER 9.0 (Case et al., 2006) at 0 K using implicit water-solvent model (igb = 5). In this program distance constraints were evaluated with an energy potential that penalised deviations from the upper limits by a $k_2|d_{max}-d_{real}|$ term. (k_2=2.0 was the force constant and d_{max} and d_{real} the upper limit and real measured distances respectively). In the case of angular restraints, a similar potential was applied on the two boundaries of the allowed range $rk_2|a_{max}-a_{real}|$ and $rk_3|a_{max}-a_{real}|$. The force constant values were determined experimentally (rk_2=rk_3=60-80). Finally, obtained structures were analysed using MOLMOL (Koradi et al., 1996) and PyMOL (www.pymol.org/) for molecule viewing and CYANA 2.1 for quality parameter calculation.

2.12.4. NMR titrations and spectra comparison

Spectra comparison was used along the thesis as a powerful tool for the exploration of changes in protein structures under different conditions: addition of binding partners (*e.g.* nucleic acids or proteins) or truncation of protein constructs. The basis of this method assumes that changes in the chemical environment of a nucleus (produced by the different conditions) produce variations on the chemical shift values of these nuclei. Therefore those signals perturbed in spectra by, for example the addition of a determined nucleic acid, correspond to amino acids that are either directly involved in the interaction or affected by the binding and the peaks that maintain their chemical shift values correspond to nuclei positioned far away from the binding interface and the putative structural changes produced by the interaction. ^{1}H-^{15}N HSQC spectra are normally used for these comparisons as it is simple to obtain, to analyse and contain backbone amide correlations, which maps all residues in the protein (except prolines).

The spectra comparison involves two steps: different spectra assignment and analysis of the chemical shift values. The assignment of the modified conditions (*e.g.* bound proteins to a ligand or truncated constructs) could be made by transferring the assignment between spectra if the changes are not dramatically large (following the titration curves) or by using the assignment methods described above. Amide peaks assigned to the same residue but

within different conditions were compared and the averaged chemical shift changes of all of them were calculated according to the formula described in Equation 2.5.

Eq. 2.5. $$\Delta\delta^{av} = \sqrt{\left((\Delta\delta_{1H})^2 + \left({}^{\Delta\delta_{15N}}/_{5}\right)^2\right) \cdot 0.5}$$

2.13. REFERENCES

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. Journal of bacteriology *62*, 293-300.

Bodenhausen, G., and Ruben, D. (1980). Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy. Chemical physics letters *69*, 185-189.

Brunschweiler, L., and Ernst, R.R. (1983). Coherence Transfer in Isotropic Mixing: Application to Proton Correlation Spectroscopy. Journal of magnetic resonance *53*, 521-528.

Case, D.A., Darden, T.A., Cheatham, T.E.I., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., and Crowley, M. (2006). AMBER 9.0 (San Francisco: University of California, San Francisco).

Clubb, R.T., Thanabal, V., and Wagner, G. (1992). A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue 1HN, 15N, and 13C' Chemical Shifts in 15N-13C-Labeled Proteins. Journal of magnetic resonance *97*, 213-217.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of biomolecular NMR *13*, 289-302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR *6*, 277-293.

Dreumont, N., and Seraphin, B. (2013). Rapid screening of yeast mutants with reporters identifies new splicing phenotypes. The FEBS journal *280*, 2712-2726.

Fadouloglou, V.E., Kokkinidis, M., and Glykos, N.M. (2008). Determination of protein oligomerization state: two approaches based on glutaraldehyde crosslinking. Anal Biochem *373*, 404-406.

Grzesiek, S., and Bax, A. (1992a). Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. Journal of the American Chemical Society *114*, 6291-6293.

Grzesiek, S., and Bax, A. (1992b). An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. Journal of magnetic resonance *99*, 201-207.

Grzesiek, S., and Bax, A. (1992c). Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein. Journal of magnetic resonance *96*, 432-440.

Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. Journal of molecular biology *273*, 283-298.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of Escherichia coli with plasmids. Gene *96*, 23-28.

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. Journal of molecular graphics *14*, 51-55, 29-32.

Kumar, A., Ernst, R.R., and Wuthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun *95*, 1-6.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Manske, M. (2006). GENtle, a free multi-purpose molecular biology tool. In Research group of Prof Dr H W Klein (Universität zu Köln).

Migneault, I., Dartiguenave, C., Bertrand, M.J., and Waldron, K.C. (2004). Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. BioTechniques *37*, 790-796, 798-802.

Milligan, J.F., and Uhlenbeck, O.C. (1989). Synthesis of small RNAs using T7 RNA polymerase. Methods in enzymology *180*, 51-62.

Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. Journal of bacteriology *119*, 736-747.

Olejniczak, E.T., Xu, R.X., and Fesik, S.W. (1992). A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. Journal of biomolecular NMR *2*, 655-659.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Puig, O., Rutz, B., Luukkonen, B.G., Kandels-Lewis, S., Bragado-Nilsson, E., and Seraphin, B. (1998). New constructs and strategies for efficient PCR-based gene manipulations in yeast. Yeast *14*, 1139-1146.

Rosales, T., and Royer, C.A. (2008). A graphical user interface for BIOEQS: a program for simulating and analyzing complex biomolecular interactions. Analytical biochemistry *381*, 270-272.

Royer, C.A., Smith, W.R., and Beechem, J.M. (1990). Analysis of binding in macromolecular complexes: a generalized numerical approach. Analytical biochemistry *191*, 287-294.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of molecular biology *189*, 113-130.

Teem, J.L., and Rosbash, M. (1983). Expression of a beta-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the rna2 mutation of yeast. Proc Natl Acad Sci U S A *80*, 4403-4407.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687-696.

Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. (1993). Two-Dimensional NMR Experiments for Correlating 13CB and *Hd/e Chemical Shifts of Aromatic Residues in 13C-Labeled Proteins via Scalar Couplings. Journal of the American Chemical Society *115*, 11054-11055.

3. Structure and function of Gbp2p and Hrb1p

3.1. INTRODUCTION

As outlined in the general introduction chapter, the first part of this work was focussed into two *Saccharomyces cerevisiae* proteins involved in mRNA transcription/export: Gbp2p and Hrb1p. They are two paralogous proteins that likely evolved from a common ancestor originated after a whole yeast genome duplication (orthologs) (Byrne and Wolfe, 2005; Dujon, 2010) and share a high degree of homology (47% identity) (Hacker and Krebber, 2004). Gbp2p and Hrb1p display a common domain architecture: a serine/arginine (SR) rich region in their N-terminus and three RRM domains (Hacker and Krebber, 2004) (Figure 3.1). At the genetic level they present one difference: Hrb1p contains an intron in its N-terminal part while Gbp2p gene is intronless (Juneau et al., 2007).

Apart from this sequence comparison, the information about Gbp2p and Hrb1p and the role they play in mRNA metabolism is scarce and somewhat fragmented. It is suspected that they may have some unknown function along transcription and mRNA export, thus they are localized together with mRNPs (Hacker and Krebber, 2004; Windgassen and Krebber, 2003; Windgassen et al., 2004). Taking into account their similarity to SR protein family in high eukaryotes (most of them alternative splicing factors), they were also related with mRNA splicing in yeast (Birney et al., 1992; Warkocki et al., 2009). In addition, these proteins are shuttled to the cytoplasm, where they have been related with post-transcriptional regulation (Mitchell et al., 2013). Finally, Gbp2p has been linked with DNA telomere maintenance in the nucleus (Hiraga et al., 2008; Pang et al., 2003). Nevertheless, none of these evidences is sufficiently strong to determine the principal biological role of these proteins.

3.1.1. Transcription, mRNA processing and export

Gbp2p and Hrb1p are defined as poly (A)+ mRNA binding proteins (Hacker and Krebber, 2004; Windgassen and Krebber, 2003) that are constituents of the mRNP particles in the nucleus and travel through the nuclear pore to the cytoplasm. Their role in formation of export-competent mRNPs is poorly understood, but some details of their interactions with other components of the mRNPs are known (Figure 3.2). Gbp2p and Hrb1p are recruited to the nascent mRNA by interactions with the THO complex (Hacker and Krebber, 2004; Hurt et al., 2004; Reed and Cheng, 2005), a large 5-subunit assembly (Tho2p, Hpr1p, Mft1p, Thp2p, and Tex1p) loaded co-transcriptionally on mRNAs, which together with other proteins (*i.e.*

Sub2p and Yra1p) form the TRanscription and EXport (TREX) complex, coupling both processes (Strasser et al., 2002). Interaction databases describe multiple hits of Gbp2p/Hrb1p to TREX proteins (acting either as baits or prey) and Mft1p and Thp2p were first proposed as Gbp2p/Hrb1p binding partners by different nuclear-cytoplasmic export requirements (Hacker and Krebber, 2004), although the direct physical interaction was not proved. This represents a difficult task because the THO forms a stable multiprotein unit (Peña et al., 2012).

Both paralogous proteins belong, together with Nlp3p, to the SR type protein family (Hacker and Krebber, 2004) (Figure 3.1). Most members of this family in higher eukaryotes usually play important roles as splicing factors, in particular in the regulation of alternative splicing (Fu, 1995; Zhou and Fu, 2013). Although Gbp2p was first described as the yeast homolog of the human splicing factor U2AF65 (Birney et al., 1992), only Npl3p has been directly implicated in splicing in yeast (Kress et al., 2008). In addition, Gbp2p and Hrb1p were detected as part of the splicing machinery (Warkocki et al., 2009), but without specifying any possible function for them, therefore the relationship between splicing and Gbp2p/Hrb1p has not been clearly established.

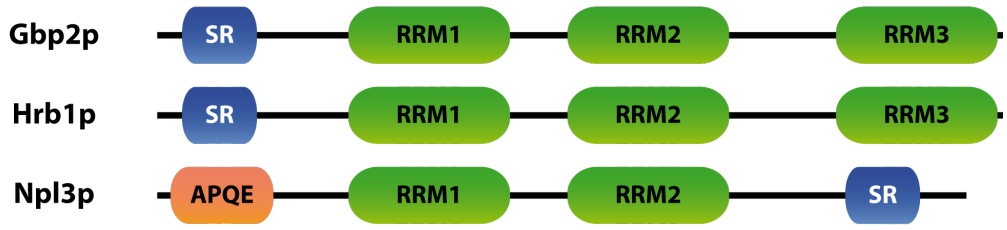


Figure 3.1. – Domain architecture of yeast SR proteins, Gbp2p, Hrb1p and Npl3p. The three proteins share an serine/arginine (SR) rich region (for phosphorylation and nuclear import) and RNA-recognition motifs (RRM). Npl3p lacks of the third RRM domain but contains an additional N-terminal region with APQE amino aci repetitions.

Interaction of both proteins with RNA (Hacker and Krebber, 2004; Windgassen and Krebber, 2003) and DNA (Lin and Zakian, 1994; Pang et al., 2003) has been demonstrated and high affinity sequence motifs have been proposed for each protein by SELEX methodology (Riordan et al., 2011). Gbp2p and Hrb1p together with the mRNP are exported to the cytoplasm through the nuclear pore and removed from the particle sometime between mRNP reorganization and the pioneer round of translation (Windgassen et al., 2004). The mRNA releasing is performed by the phosphorylation of the SR region: the cytoplasmic kinase Sky1p is the responsible for this reaction in Gbp2p (as in Npl3p) (Gilbert et al., 2001; Windgassen

and Krebber, 2003), but Hrb1p phosphorylation is not clear and could be performed by Sky1p action or by another SR protein kinase yet to be discovered (Porat et al., 2006). Finally, the import to the nucleus is carried out by the importin Mtr10p (Senger et al., 1998; Windgassen and Krebber, 2003; Windgassen et al., 2004).

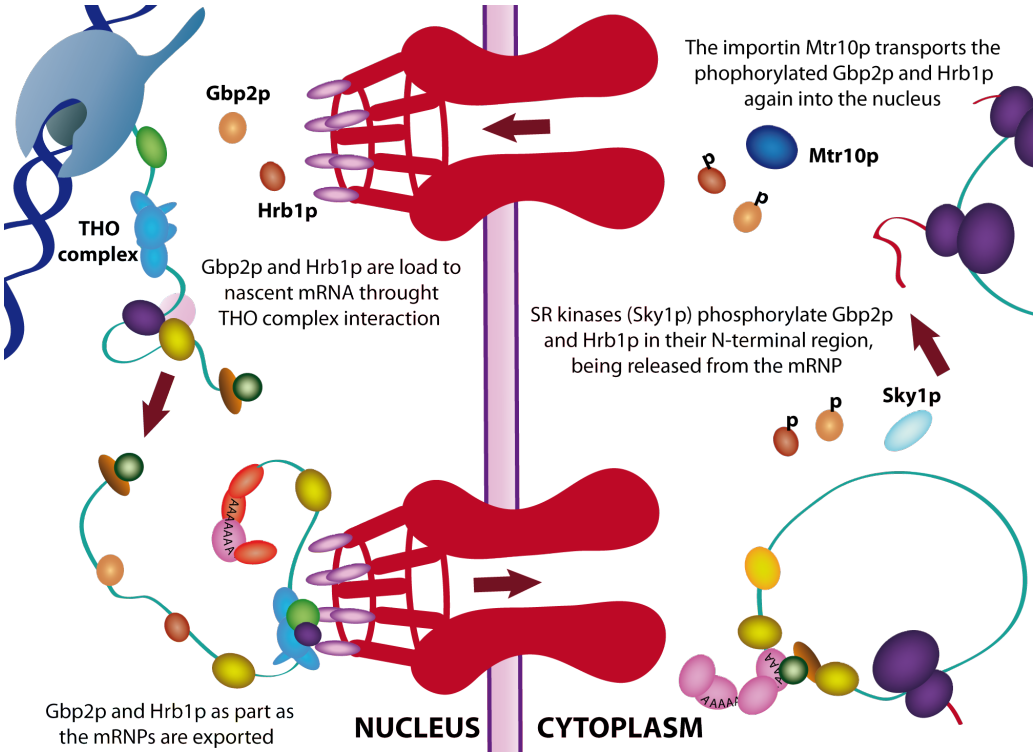


Figure 3.2. – Representation of Gbp2p and Hrb1p along transcription and export processes.

Despite this knowledge, the functions played by Gbp2p and Hrb1p in mRNP biogenesis are unclear; null mutant strains do not present growth phenotypes and mRNAs seem to be formed and delivered correctly in them (Giaever et al., 2002). In contrast, overexpression of Gbp2p is cytotoxic (Pang et al., 2003).

3.1.2. Post-transcriptional control

As it is further described in the chapter 5 of this thesis, post-transcriptional control is a powerful mechanism of gene regulation that is strongly correlated to respond to different environmental stresses. In the cytoplasm, mRNAs are grouped together with other proteins and partially assembled ribosomes to form granular structures in which the translation is

arrested. There are several different granules with variable composition depending on the nature of the environmental stress. Gbp2p and Hrb1p were detected in two of these structures: stress granules (Gbp2p) and in P-bodies (Hrb1p) (Buchan et al., 2008; Mitchell et al., 2013). But this localization (as in the previous cases) does not imply an assignment of the role they may play in these structures and so far neither of these proteins has been regarded as essential for the formation of these cytoplasmic granules.

3.1.3. Implication of Gbp2p in telomere maintaining

In a screening of yeast protein libraries, Gbp2p was shown to be capable to bind to a telomeric ssDNA (Lin and Zakian, 1994). The protein was identified as a G-strand DNA binding protein, which presented high affinity for yeast telomeric DNA sequences. However subsequent works reveal that Gbp2p is not directly involved either in telomere length control, or in telomere positioning (TPE), or in telomere transcription repression (Konkel et al., 1995).

After years of silence, recently, some genetic interactions with two essential proteins in telomers (Cdc13p and Rap1p) were detected (Hiraga et al., 2008; Pang et al., 2003), reviving the interest about the possible role of Gbp2p in telomere biochemistry. In the first case, the Cdc13p mutant phenotype, that causes accumulation of single stranded telomeric DNA near telomeres at high temperatures, was restored by overexpression of Gbp2p, probably by the ssDNA covering by Gbp2p excess (Pang et al., 2003). And deletion of Gbp2p alters Rap1p localization (an essential protein for telomere length control, telomere positioning and telomere transcription repression (Pina et al., 2003)): in wild type cells Rap1p forms granules together with telomers in the nuclear periphery. Whereas in Gbp2p mutants Rap1p still organizes in granules (conserving its functions), but these granules are not near to the nuclear envelopment (Hiraga et al., 2008). Surprisingly, the relation with telomeric DNA and these two genetic interactions were not detected using the homolog protein Hrb1p (Hiraga et al., 2008; Konkel et al., 1995).

As in the previous cases, the linkage between Gbp2p and telomers is unclear and moreover, the relationships between this possible Gbp2p function and others in transcription, splicing, export or in translational control are totally unknown.

3.1.4. Objectives

Taking into account the scarce and variable information published about Gbp2p and Hrb1p, the study of these two proteins was focussed into the analysis of the differences between their three RNA recognition motifs (RRMs), from chemical physical, structural and functional points of view. The obtained data will be compared with those available for the other SR family members.

This study involved four steps. A first step of protein characterization using different biochemical and chemical-physical approaches, a second step based on NMR structural study, a third step in which the nucleic acid binding interaction of the different motifs was characterized and a final step for exploring the global function of Gbp2p and Hrb1p and the specific role of their domains.

3.2. RESULTS

3.2.1. Bioinformatic study and design of protein constructs

The use of bioinformatic tools in the earliest steps of biophysical and structural protein studies has became a mandatory practice during last years. This has been boosted by the exponential growth of sequence databases. Bioinformatics provide useful information about potential new protein targets, by predicting their structural organization (domains), secondary structure, sequence conservation, possible posttranslational modifications and even the existence of intrinsically unstructured regions. The traditional approach followed in protein research is the "divide and conquer" strategy in which large and structurally heterogeneous proteins are split into small and tractable constructs that nevertheless retain the same structure information. This approach minimizes common problems in protein studies such as aggregation or instability and facilitates the work in those techniques where large protein size and/or conformational heterogeneity represent limiting factors (*e.g.* NMR and X-ray crystallography). The outcome of modern bioinformatics tools in combination with ever-growing databases has converted the "ancient art" of choosing the right domain boundaries into an easier and more rational task.

Gbp2p and Hrb1p have a similar domain organization (Figure 3.3): an intrinsically unstructured N-terminal region and three RNA recognition motifs (RRM). The N-terminal part

is predicted to contain some serine phosphorylation sites, probably involved in their nuclear import regulation (Hacker and Krebber, 2004; Windgassen and Krebber, 2003). The three RRMs are unevenly distributed along the protein: RRM1 and RRM2 appear in tandem, separated by a segment of about 20 residues and RRM3 is placed just at the C-terminus of the protein, separated from RRM2 by around 35 (Hrb1p) or 50 (Gbp2p) amino acids. Taking this information into account, six different constructs were designed for each protein, all of them including between one and three RRM domains: RRM1, RRM2, RRM3, RRM12, RRM23 and RRM123 (Figure 3.3).

Figure 3.3. – Schematic view of Gbp2p and Hrb1p domain organization, SR indicates the serine-arginine rich region and RRMX each RNA Recognition Motive and the representation of the six different proposed constructs.

Two sequence-based analysis were performed to define the domain boundaries of each construct precisely: a secondary structure prediction (details in Appendix 2) and a sequence alignment of homologous proteins to Gbp2p and Hrb1p (Appendix 3).

As expected, structural prediction identified three regions on both proteins that roughly coincide with the three proposed RRMs and include the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ arrangement. Further, potential coiled coil or α-helical region of about ten residues is proposed after RRM2 in Gbp2p and a short α-helix was identified just before RRM3 in Hrb1p (Figure 3.4). Finally, the N-terminal region seems to be unstructured in both proteins.

The sequence alignment of Gbp2p and Hrb1p related proteins was made by comparison of a list of putative fungal homologs, after excluding those from the *Saccharomyces* genus which show a very high degree of identity. The result of the alignment (Appendix 3) identified three conserved regions centred around each RRM domain, boundaries of RRM1 and RRM2 matched almost exactly with the predictions made by automatic domain search engines of the Uniprot and other databases. In contrast, high conservation was found for a short sequence just preceding RRM3, which coincides with the α-helix propensity region detected in Hrb1p (Figure 3.5). In addition, a highly conserved serine/arginine rich region was detected in the N-

terminal part of both proteins and some residues just preceding RRM1 and in the RRM1-RRM2 linker were also conserved.

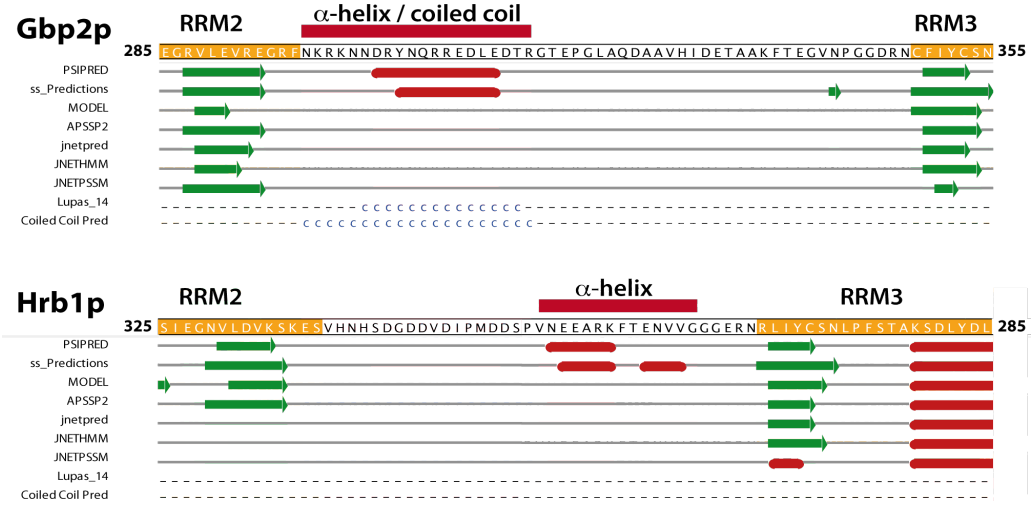


Figure 3.4. – Secondary structure prediction for the linker between RRM2 and RRM3 of Gbp2p (up) and Hrb1p (down) proteins, showing the presence of putative helical elements. The full-sequence predictions are shown in Appendix 2.

The final choice of the construct boundaries was made considering all of these observations and is summarized on Figure 3.6. The main differences with the predicted domain boundaries in the database are the incorporation of ~20 amino acids in the N-terminal part of RRM1 and RRM3 and another 20 residues after the predicted C-termini of RRM2. Both regions were incorporated to Gbp2p and Hrb1p protein constructs.

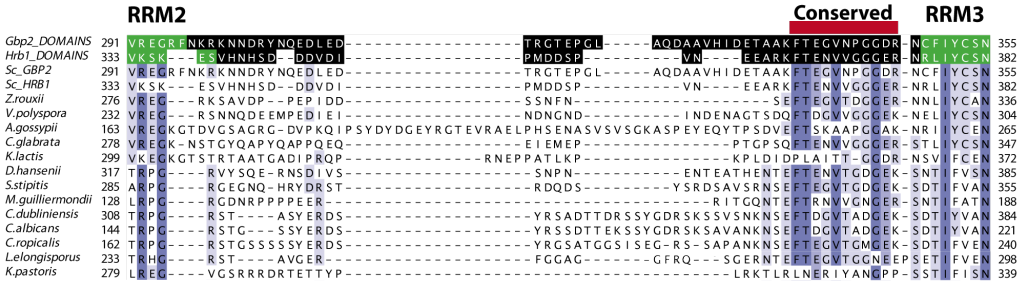


Figure 3.5. – Gbp2p and Hrb1p homologous alignment in the region between RRM2 and RRM3 (domains are in green and linker sequence in black), highlighting the conservation of the N-terminal region before RRM3. Full-sequence alignment is included in Appendix 3.

Premature truncation of structural domains can lead to misfolding or a loss of activity. Therefore, the protein dissection approach requires tools for quality control in order to validate whether the generated constructs are structural and functional representatives of the whole protein. Measurement of an enzymatic activity and other physical-chemical properties of the constructs are frequently used as validation tools. The following experiments for Gbp2p and Hrb1p aim to validate the proposed constructs for further structural and functional analyses and at the same time to provide useful low-resolution structural information.

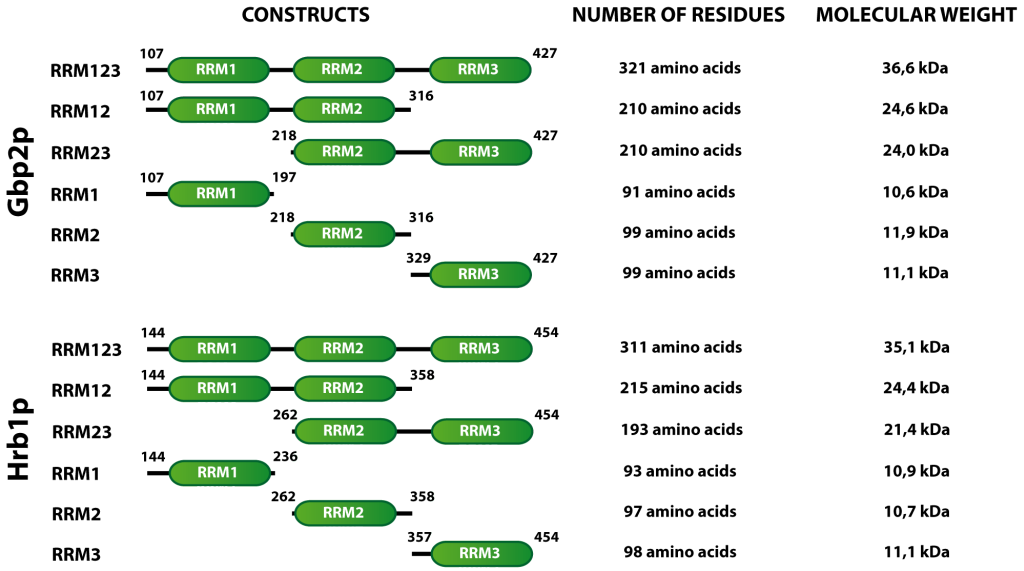


Figure 3.6. – Final boundaries of each Gbp2p and Hrb1p construct.

3.2.2. Protein expression and purification

All the constructs were obtained following similar protocols (see 2.2 section for details). These are based on bacterial (*E.coli*) over-expression, which is widely used because of its robustness, simplicity of cloning process and adaptability to produce isotope-labelled protein samples. Gbp2p and Hrb1p constructs (Figure 3.6) were cloned into a home modified vector with a N-terminal thioredoxine A fusion tag (to enhance expression and solubility), a 6 x histidine peptide (to simplify purification procedure) and a TEV site (ENLYFQ/GS) placed between the fusion tag and the target proteins (for final release of the target product). The total length of this N-terminal fusion tag is 133 amino acids (14.4 kDa).

Preliminary tests on Gbp2p and Hrb1p constructs showed that all expressed as soluble proteins, with good yields and apparently without degradation (Figure 3.7). Therefore preparative protein expression of all Gbp2p and Hrb1p constructs was carried out.

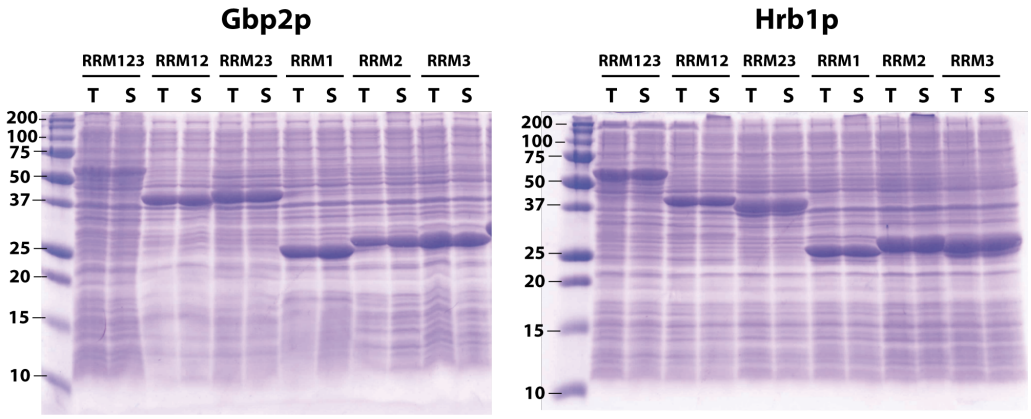


Figure 3.7. – SDS-PAGE of expression tests of all Gbp2p and Hrb1p constructs. Two lanes, T (total protein fraction) and S (soluble proteins) were included for each construct. Theoretical molecular weight of each fusion protein: Gbp2p – RRM123 50.9 kDa, RRM12 38.8 kDa, RRM23 38.2 kDa, RRM1 24.9 kDa, RRM2 26.2 kDa and RRM3 25.3 kDa; Hrb1p – RRM123 49.4 kDa, RRM12 38.6 kDa, RRM23 35.7 kDa, RRM1 25.2 kDa, RRM2 24.9 kDa and RRM3 25.4 kDa.

Cell lysis and a purification process, which involved several chromatography steps including a TEV digestion to remove the protein tag, were done to get the protein constructs at preparative level. They were obtained with excellent yields and a high degree of purity. The exceptions were the constructs starting at RRM2 domain (Gbp2p RRM2, Gbp2p RRM23, Hrb1p RRM2 and Hrb1p RRM23) for which TEV digestion was less effective (Figure 3.8.A). This minor problem, which was overcome by adding more protease, suggests that the TEV target sequence is less accessible in these constructs. Further problems were found for Gbp2p RRM23 construct during ion exchange chromatography step as it precipitated inside the column and could only be recovered under denaturing conditions (8 M urea buffers).

Remarkably, the presence of different types of reduction agents in the composition of the purification buffers affects sample homogeneity, likely due to the reactivity of some exposed cysteines. This effect was studied in detail concluding that the use of β-mercaptoethanol during the purification leaded to irreversible covalent modification of the recombinant proteins (Appendix 4). This chemical probably forms disulphide-bridged adducts with target cysteines, which are highly resistant to reductive conditions. Similar protein modifications

have been reported in the literature (Fujiwara et al., 2007; Ihara et al., 2012; Zhukov et al., 2008) and are used to stabilize protein products for pharmaceutical applications. An optimal protein purification protocol using high concentrations of other reduction agents (DDT and TCEP) was followed to avoid sample heterogeneity.

Excluding Gbp2p RRM23, all the protein constructs were easily purified and could be prepared at high concentration (up to 500-1000 μM, depending of the construct). SDS-PAGE (Figure 3.8.B), NMR and mass spectrometry analyses confirmed high purity and homogeneity of the samples. In addition, samples could be frozen on liquid nitrogen and stored at -20ºC safely without apparent loss of properties or structure.

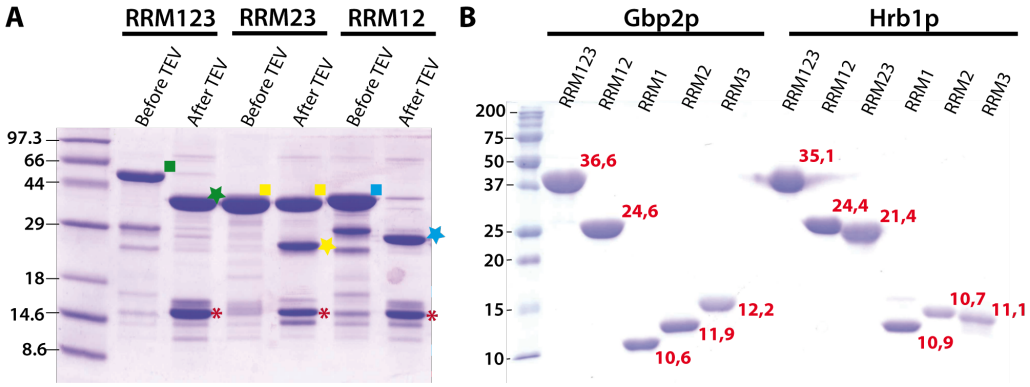


Figure 3.8. – A) SDS-PAGE of TEV digestion of some Hrb1p constructs. Two lanes for each construct (green for RRM123, yellow for RRM23 and blue for RRM12) show the fusion protein before TEV treatment (square) and the resultant proteins after digestion (a star marks the different protein constructs and a red asterisk the digested thioredoxine A tag). RRM123 and RRM12 present almost total digestion, but RRM23 shows a loss of TEV effectiveness. B) SDS-PAGE of all purified samples of each Gbp2p and Hrb1p construct. Protein ladder is in left lane and each band is marked with the corresponding molecular weight (kDa).

3.2.3. Physical-chemical characterization

The purified protein constructs were studied by various physical-chemical methods (*i.e.* circular dichroism and fluorescence spectroscopies, gel filtration chromatography and NMR) in order to obtain relevant information about their stability, secondary structure content, oligomeric state, etc. The results were useful to determine which constructs behave as folded and monodisperse samples and hence are good candidates for further structural and functional studies, and which ones are not suitable due to construct truncation, misfolding, aggregation or other problems.

3.2.3.1. Analytical gel filtration chromatography

Gel filtration chromatography can be used with either preparative or analytical applications. In the analytical one, the chromatogram provides information about the apparent molecular weight of the studied protein (oligomerization state) and the different species present in the sample (level of homogeneity).

Samples of different Hrb1p and Gbp2p protein constructs were loaded into a Superdex 200 column at concentrations of 100 μM approximately. This value coincides with the lower limit of the concentration ranges used in NMR analysis (100 μM to 1mM), therefore the observations are useful to interpret the NMR data as well. Gel filtration chromatograms of Gbp2p and Hrb1p constructs display a typical single-peak profile and, in general, each peak appears at exclusion volumes close to the expected values according to their molecular weights (Figure 3.9).

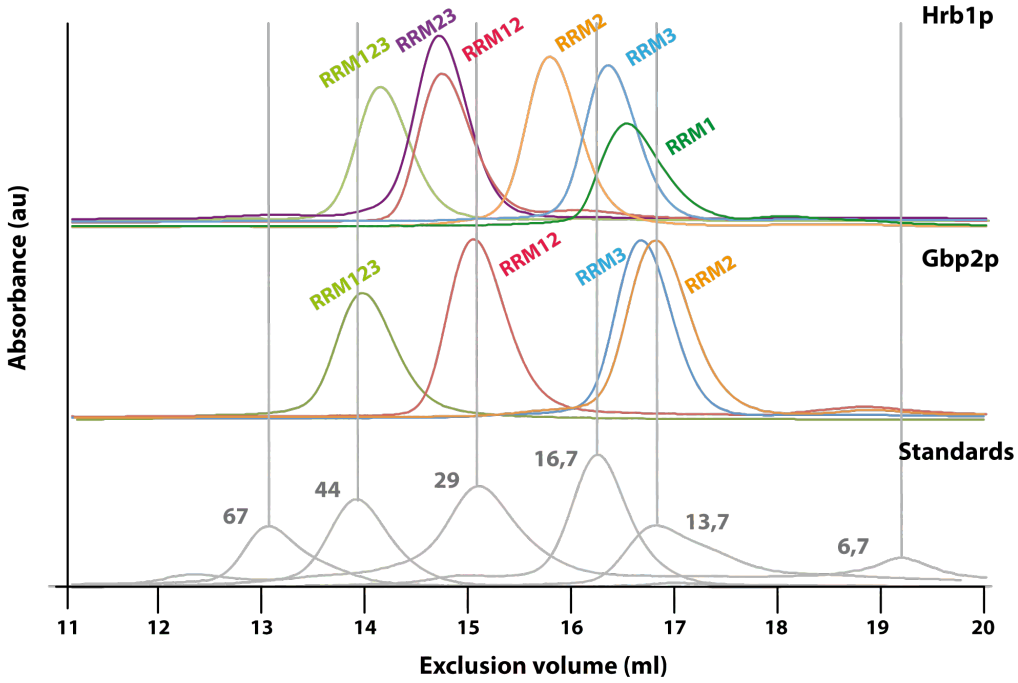


Figure 3.9. – Superdex 200 gel filtration chromatograms of the different Hrb1p (up) and Gbp2p (middle) constructions compared with a set of standards (down) (see 2.8 section). The corresponding molecular weights are: Hrb1p – RRM123 35.1 kDa, RRM12 24.4 kDa, RRM23 21.4 kDa, RRM1 10.9 kDa, RRM2 10.7 kDa, RRM3 11.1 kDa; Gbp2p – RRM123 36.6 kDa, RRM12 24.6 kDa, RRM2 11.9 kDa, RRM3 12.2 kDa.

Hrb1p RRM2 elution chromatogram shows a different behaviour, as its peak appears earlier than other Gbp2p/Hrb1p single RRMs but later than RRM tandems, indicating that this construct has a bigger hydrodynamic radius than other RRMs. This fact could be caused either by the existence of an oligomerization equilibrium or alternatively by the contribution of partially unfolded regions. However this odd behaviour disappears in the Hrb1p RRM12 tandem suggesting that the effects are somewhat compensated in this construct.

In summary, the gel filtration data suggest that Gbp2p and Hrb1p constructs, with the exception of Hrb1p RRM2, are *a priori* good candidates for further structural studies by NMR.

3.2.3.2. Fluorescence spectroscopy

Aromatic amino acids contribute to intrinsic fluorescence in proteins and this effect is additive and dominated by the number of tryptophans in the sequence. This aromatic side-chain has a maximum of absorbance at 280 nm and the emission spectrum maximum ranges from 300 to 350 nm depending on its local environment. Water-exposed tryptophans emit at a maximum at 352 nm, whereas tryptophans situated in protein hydrophobic pockets emit at lower wavelengths (around 330 nm) (Royer, 1995). This characteristic makes tryptophan residue an excellent intrinsic probe in the study of proteins, either as a simple way to observe the environment of this residue (if it is exposed or buried) or as a reporter to follow protein denaturation processes or molecular interactions.

Gbp2p and Hrb1p contain only one tryptophan that, according to secondary structure predictions, is located in the α_1-helix of RRM2. Experimentally, samples of tryptophan-containing constructs (RRM123, RRM12, RRM23 and RRM2) were prepared at a concentration equivalent to an absorbance of 0.03 uA at 280 nm and fluorescence emission between 300 – 450 nm was recorded using an excitation wavelength of 290 nm to avoid tyrosine and phenylalanine excitation. For all cases, the emission spectra presented a maximum centred around 350 nm (Figure 3.10). These results clearly showed that the tryptophan is solvent exposed in all constructs of both proteins. Thermal denaturing experiments were recorded but, as expected, the spectra obtained did not show remarkable differences in the fluorescence emission maximum, reflecting that the tryptophan local environment does not change upon denaturalization.

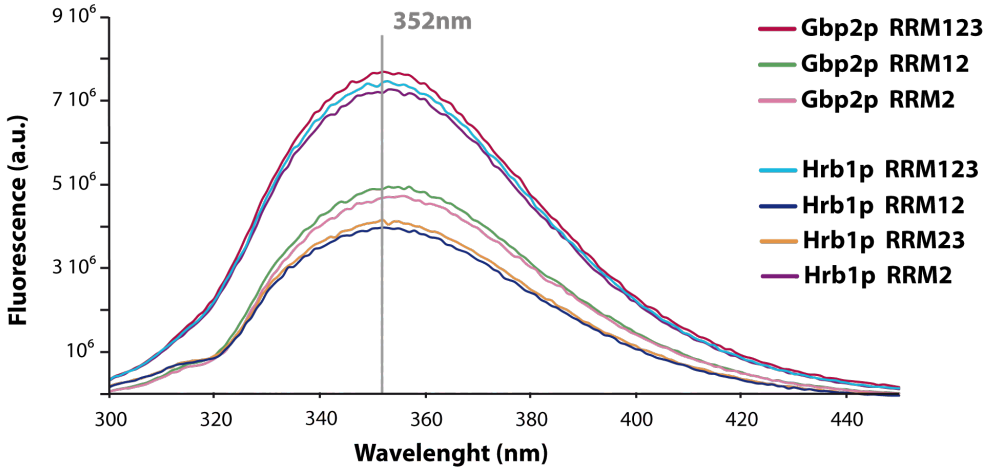


Figure 3.10. – Emission spectra of the Gbp2p and Hrb1p constructs including RRM2 domain. The maximum in all cases is placed around 352 nm, consistent with solvent-exposed tryptophan environment for all constructs.

3.2.3.3. Circular dichroism spectroscopy

Far UV circular dichroism spectroscopy is a useful technique for protein secondary structure determination because the measured variable, ellipticity, depends basically on the peptide bond angles ϕ and ψ, which are different in each secondary conformation. For that reason, they present different spectrum (Figure 3.11.A): α-helical components typically produce a maximum of ellipticity around 190 nm and two minimums at 208 and 225 nm, β-sheet conformations present more variability and less intense spectra and they usually produce an ellipticity maximum at 195 nm and a minimum at 210-220 nm, and unstructured or random coil conformations present a minimum of ellipticity at about 195 nm and a small maximum at 220 nm (Ranjbar and Gill, 2009).

For the study of Gbp2p and Hrb1p constructs, far UV CD spectra of each one was acquired at 20ºC using samples prepared at a concentration of 0.2 mg/ml (Figure 3.11.B). Unfortunately, equipment limitations avoided the acquisition at wavelengths lower than 200 nm with good signal to noise ratio. All the CD spectra of the constructs were quite similar and presented a minimum at about 208 nm and a small trough around 220 nm (the intensity of this minimum varied between different constructs). These characteristics correspond to the presence of α-helical and β-sheet secondary structure elements (the canonical RRM domain contains these two types).

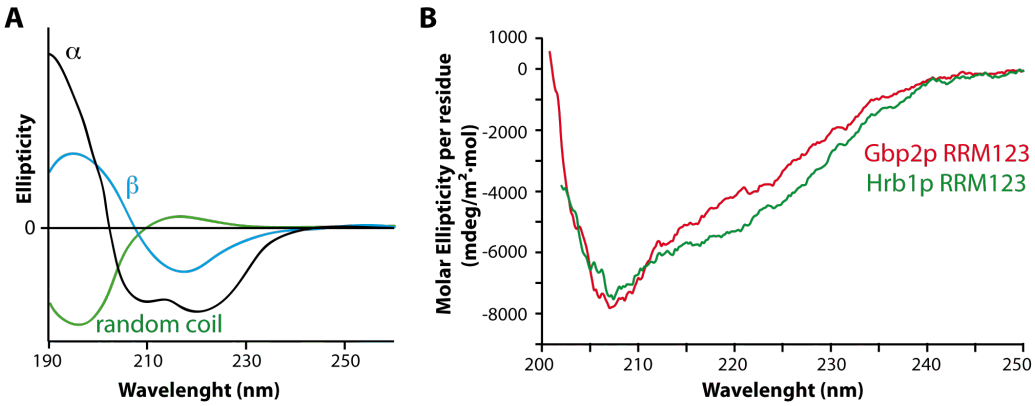


Figure 3.11. – A) Representation of theoretical α-helix (black), β-strand (blue) and random coil (green) circular dichroism spectra. B) Far UV circular dichroism spectra of Gbp2p RRM123 construct (red) and Hrb1p RRM123 one (green).

In addition, some circular dichroism experiments were used to study denaturation process. First, an exploratory set of spectra was acquired at different temperatures (20, 35, 50, 70 and 90ºC) for each construct. Except Gbp2p RRM2, all the constructs showed major variations at wavelengths lower than 210 nm (Figure 3.12): the α and β maximum at 195 – 200 nm gradually turned into the random coil minimum due to disruption of the native structure. Besides, other slight changes appeared in some constructs in the range of 210 – 250 nm, also attributable to the secondary structure loss. In contrast, the CD spectra of the Gbp2p RRM2 construct did not evolve towards the random coil minimum during melting, rather great changes at around 215 – 225 nm were seen (Figure 3.12) and in addition, the protein precipitates upon heating.

After examining these exploratory spectra and considering the hardware limitations, a wavelength of 203 nm was selected to monitor the process of protein denaturation *vs* temperature. Ellipticity was recorded at this wavelength from 20 to 90 ºC for samples of all constructs (Figure 3.12), and additionally at 217 nm for Gbp2p RRM2 construct. The circular dichroism signals were also monitored from 90 to 20ºC to check reversibility (renaturation).

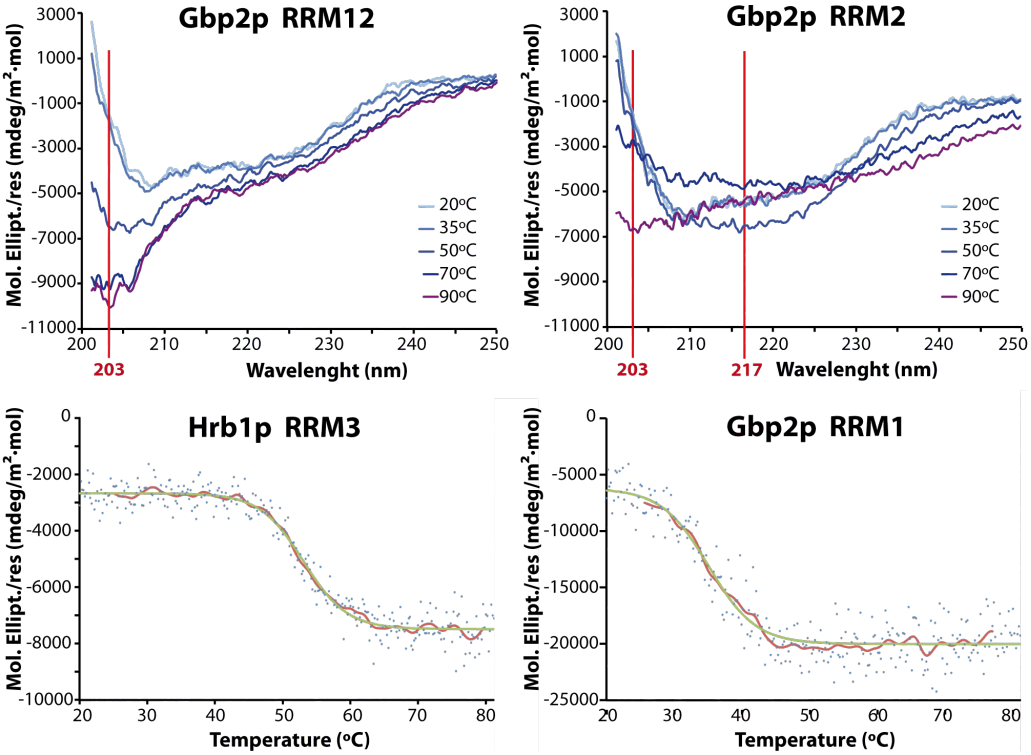


Figure 3.12. – Up) Far UV CD spectra of two constructs recorded at 20, 35, 50, 70 and 90 ºC. Protein denaturation process is evidenced by the different spectra shape at each temperature. Down) Denaturation curves of two Gbp2p and Hrb1p constructs, the recorded data are represented by blue points, the red line represents the smooth data and the fitting to a Boltzmann sigmoid is in green. Equivalent data for all constructs are provided in Appendix 5.

Unfortunately, quantitative renaturation could not be achieved in any case (Table 3.1), only Hrb1p RRM12 and Hrb1p RRM1 constructs could recover about 80% of native ellipticity after the slow cooling. The lack of reversibility precludes a rigorous thermodynamical analysis, thus apparent melting temperatures ($T_{1/2}$) were derived for each protein (except for Gbp2p RRM2) by fitting the data to a two states Boltzmann sigmoid (Equation 3.1).

Eq. 3.1.

$$[\theta] = \frac{[\theta]_{max} - [\theta]_{min}}{1 + e^{(T - T_{1/2})/A}} + [\theta]_{min}$$

In the equation, [θ] represents the molar ellipticity at each temperature (T), $[\theta]_{min}$ and $[\theta]_{max}$ are the ellipticity values of each plateau (the denatured and folded one, respectively) and A is the slop factor that gives an idea of the cooperativity of the denaturing process. The calculated values of $T_{1/2}$ are shown in Table 3.1 and the normalized representation of each fit can be seen in Figure 3.13.

In general, Hrb1p constructs are systematically more stable than their Gbp2p counterparts. The RRM3 domains are the most stable among the single-domain construct series and Hrb1p RRM3 is the most stable construct of this protein (Table 3.1). These single RRM3 domains seem to unfold in a highly cooperative manner as denoted by the narrow steepness of the curves and their higher $T_{1/2}$ values point to the presence of additional structural elements stabilizing the canonical RRM fold. This hypothesis was later confirmed by the structures of these domains (see 3.2.4 section). Hrb1p RRM1 and RRM2 display a curve profile similar to RRM3 constructs but with lower apparent $T_{1/2}$. Gbp2p RRM1 exhibits the lowest $T_{1/2}$ among the single RRM constructs dataset, indicating that it is quite unstable when isolated. On the other hand, Gbp2p RRM2 precipitates along denaturation, having a more complex CD spectra (Figure 3.12 and Appendix 5).

Protein Construct	Apparent $T_{1/2}$ (ºC)	Renaturation (%)
Gbp2p RRM123	47.2 ± 0.3	19.8 ± 3.2
Gbp2p RRM12	45.6 ± 0.3	24.6 ± 8.7
Gbp2p RRM1	33.9 ± 0.7	< 5.0
Gbp2p RRM2	-	-
Gbp2p RRM3	49.7 ± 0.5	< 5.0
Hrb1p RRM123	51.4 ± 0.4	40.3 ± 3.1
Hrb1p RRM12	53.1 ± 0.4	82.2 ± 2.0
Hrb1p RRM23	46.7 ± 1.0	18.9 ± 16.8
Hrb1p RRM1	47.5 ± 0.4	83.6 ± 1.7
Hrb1p RRM2	41.2 ± 0.7	36.6 ± 7.4
Hrb1p RRM3	54.4 ± 0.4	< 5.0

Table 3.1. – Apparent $T_{1/2}$ values calculated for each construct by fitting the data to a Boltzmann sigmoidal curve, and percentage of signal recovered after renaturation.

Hrb1p RRM12 and Gbp2p RRM12 tandems do not recall the behaviour of their individual RRMs. This is clearly evident in the case of Gbp2p tandem, which follows an apparently conventional temperature melting process very different from the Gbp2p RRM2 one. The fact that the melting curve of Hrb1p RRM12 is well above of those of the isolated RRM domains is a strong evidence that its structure is not a simple combination of its subunits. These results suggest the presence of an additional stabilization in the tandem construct.

Finally RRM123 constructs show similar curves to those of RRM12, reflecting a dominance of this tandem in the CD spectrum over the RRM3. The inclusion of this later domain causes little

but opposite effects in both proteins. In Gbp2p, it seems to increase the stability (brown curve over green on Figure 3.13) and in Hrb1p to decrease it.

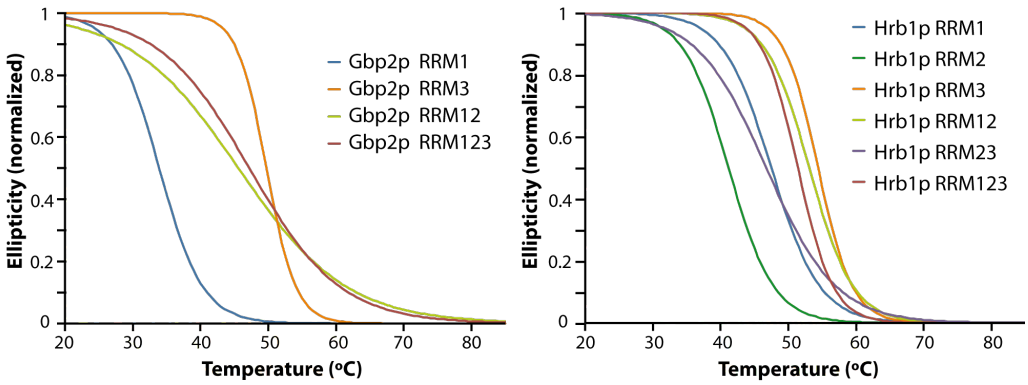


Figure 3.13. – Representation of fitting curves for each Gbp2p and Hrb1p construct.

In summary, these stability studies suggest that Gbp2p RRM123 and Hrb1p RRM123 constructs contain two well-defined structural regions that could be divided into two independent constructs: a tandem of RRM (RRM12) and a single RRM (RRM3). A further dissection of the RRM12 tandem into its individual RRM domains entails loss of stability and an odd behaviour in the case of Gbp2p RRM2. Despite their similar domain organization, Gbp2p and Hrb1p are very different from the stability point of view, in particular regarding their RRM12 tandems.

3.2.3.4. Nuclear Magnetic Resonance analysis

Finally, Gbp2p and Hrb1p protein constructs were also analysed by Nuclear Magnetic Resonance (NMR). In particular, ^{1}H-^{15}N HSQC spectra were obtained for all the protein constructs under strictly controlled experimental conditions (*i.e.* concentration, buffer composition and temperature). This type of experiment is considered the protein "fingerprint" spectrum, where each amino acid (except proline) is represented by at least one peak. The ^{1}H-^{15}N HSQC offers in general good signal dispersion without excessive peak overlapping for well-folded proteins up to 200 amino acids and is very useful to characterize molecular interactions. However, it is less useful for intrinsically unstructured proteins due to high degree of overlapping on these cases.

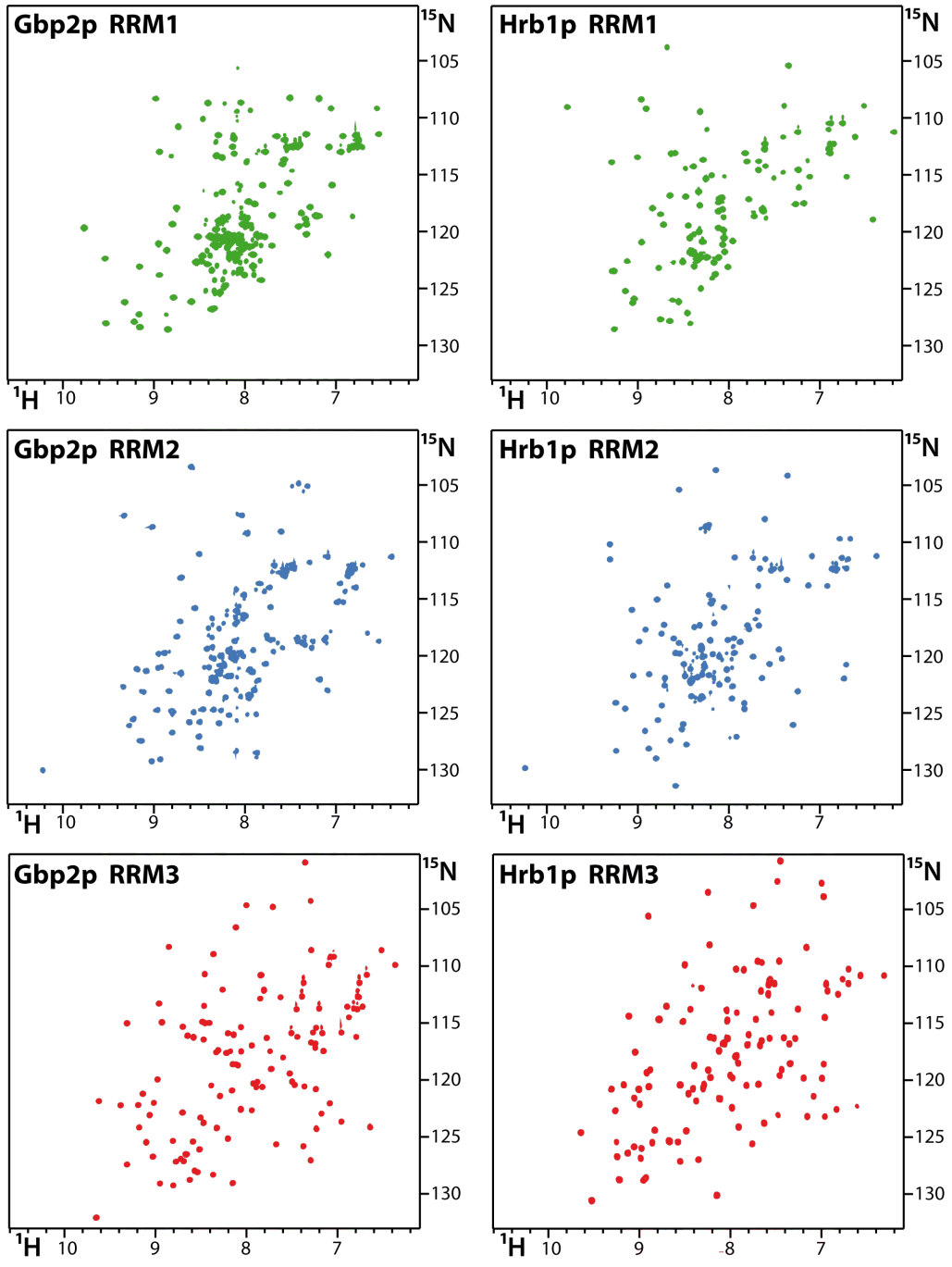


Figure 3.14. – ^{1}H-^{15}N HSQC spectra of individual RRMs of Hrb1p and Gbp2p proteins. All the spectra has been recorded in a 800 MHz spectrometer at 298 K and the constructs were dissolved in a 25 mM potassium phosphate at pH 6.5, 25 mM NaCl and 0.1 mM DTT buffer.

To ensure identical experimental conditions, ^{15}N isotopically labelled samples of all constructs were dialyzed against the same buffer, at final protein concentrations of 100-500 μM. The ^{1}H-^{15}N HSQC spectra of each construct were acquired on the same spectrometer (Bruker AV800) and at the same temperature (298 K). In general, the spectra show high degree of peak dispersion (Figure 3.14-16), which confirms the presence of well-structured domains, in consistency with circular dichroism data and the bioinformatic analysis.

The quality of the spectra is not correlated with the molecular weight of the constructs. Gbp2p RRM3 and Hrb1p RRM3 ^{1}H-^{15}N HSQC spectra present homogeneous, narrow and intense peaks while the other single-RRM spectra (Gbp2p RRM1, Gbp2p RRM2, Hrb1p RRM1 and Hrb1p RRM2) have poorer quality and heterogeneous peak intensities (Figure 3.14).

The ^{1}H-^{15}N HSQC spectra of the two RRM12 tandems are of poorer quality than expected for molecules of their sizes (20 KDa). The spectra present differential line broadening for a large subset of signals, which is more dramatic in the case of Gbp2p RRM12 where many peaks almost disappeared (Figure 3.15). This effect is attributable to chemical exchange processes but it is difficult to identify the origin of it: intramolecular or intermolecular interactions, larger conformational changes or coexistence of folding intermediates and/or unfolded species. Fortunatelly, the quality of the spectra of Hrb1p RRM12 is sufficiently good to be assigned by using triple resonance methods (see 3.2.4 section). In contrast, the spectra of Gbp2p RRM12 could only be partially assigned by comparison with the assignment of the single RRM domains.

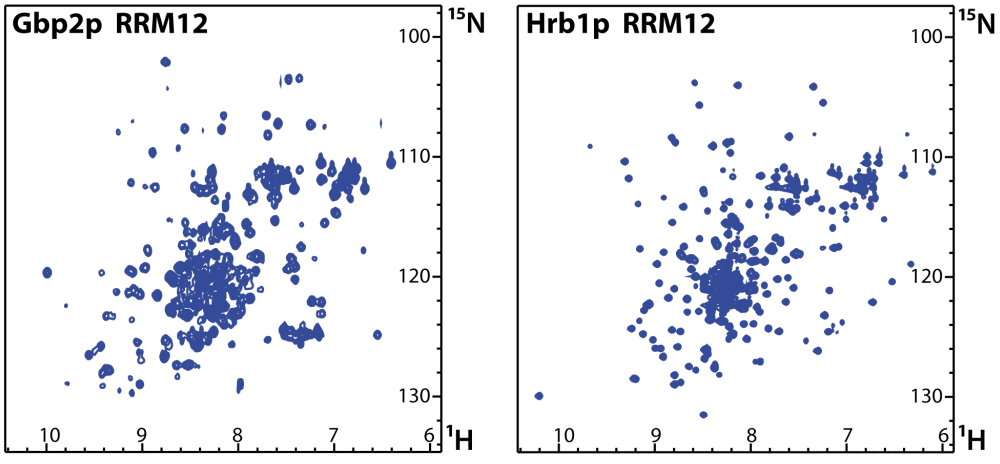


Figure 3.15. – ^{1}H-^{15}N HSQC spectra of Hrb1p RRM12 and Gbp2p RRM12 constructs recorded as identical conditions than those in Figure 3.14.

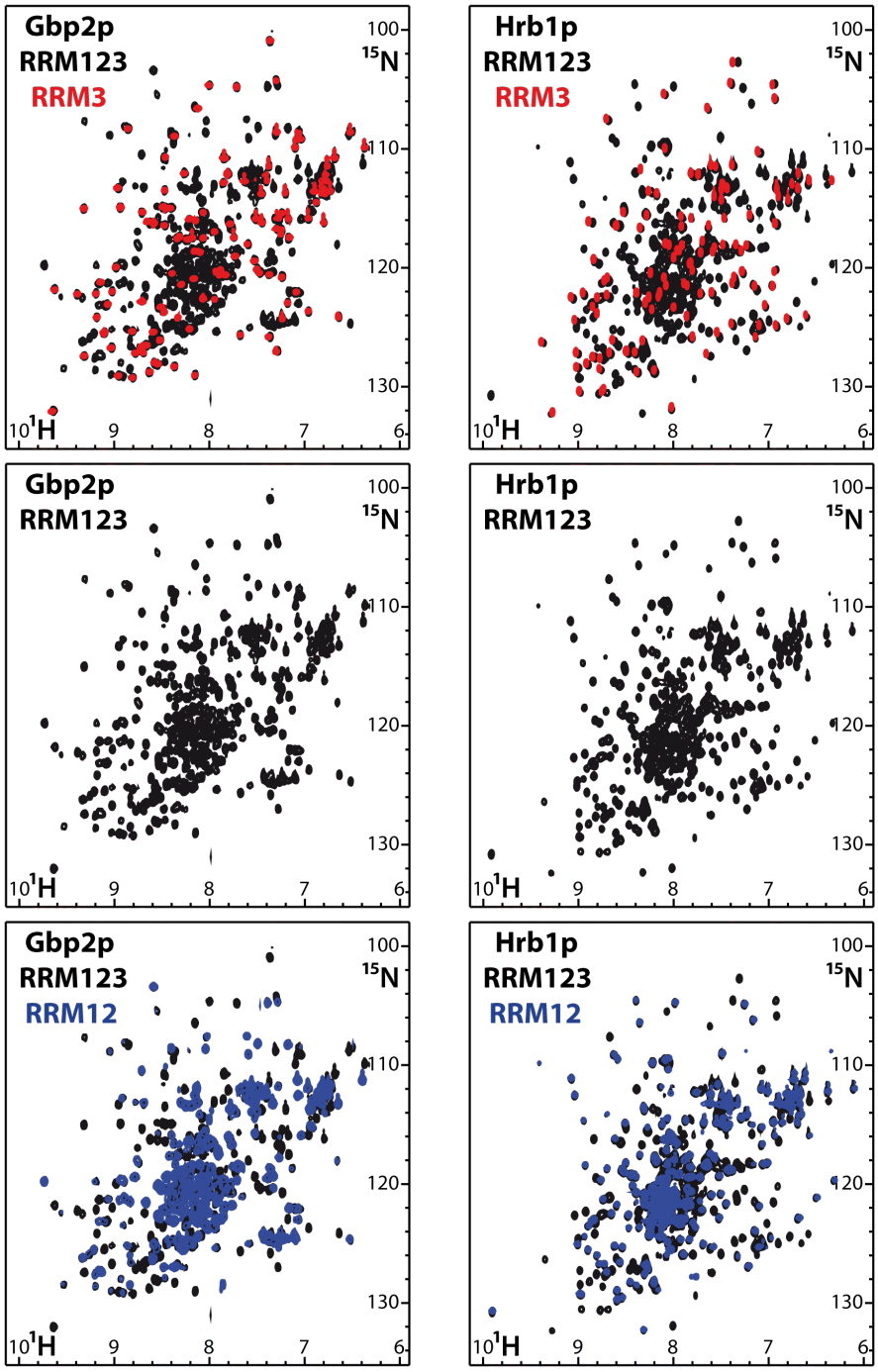


Figure 3.16. – ^{1}H-^{15}N HSQC spectra of Hrb1p RRM123 and Gbp2p RRM123 constructs (in black) compared with RRM3 constructs (red spectra upper panels) and RRM12 tandems (blue spectra in lower panels).

Finally the ^{1}H-^{15}N HSQC spectra of the two larges constructs (Gbp2p RRM123 and Hrb1p RRM123) display a coexistence of two types of well-differentiated signals: some with narrow lines and others with substantial line broadening (Figure 3.16). These two sets of signals match with the spectra of RRM3 and RRM12 respectively, indicating that these two regions behave as structurally independent subunits within the same polypeptide.

Interestingly, the spectra of RRM1 and RRM2 domains do not show the same level of similarity to RRM123 (or RRM12) than the above. Although most of the single-RRM peaks are at similar positions than in its larger constructs, some of them are certainly shifted to an extend that makes impossible to assign the spectra of the larger form by simple comparison (Figure 3.17). A plausible explanation for these observations is that RRM1 and RRM2 domains may interact each other (or with connecting elements) in the tandems.

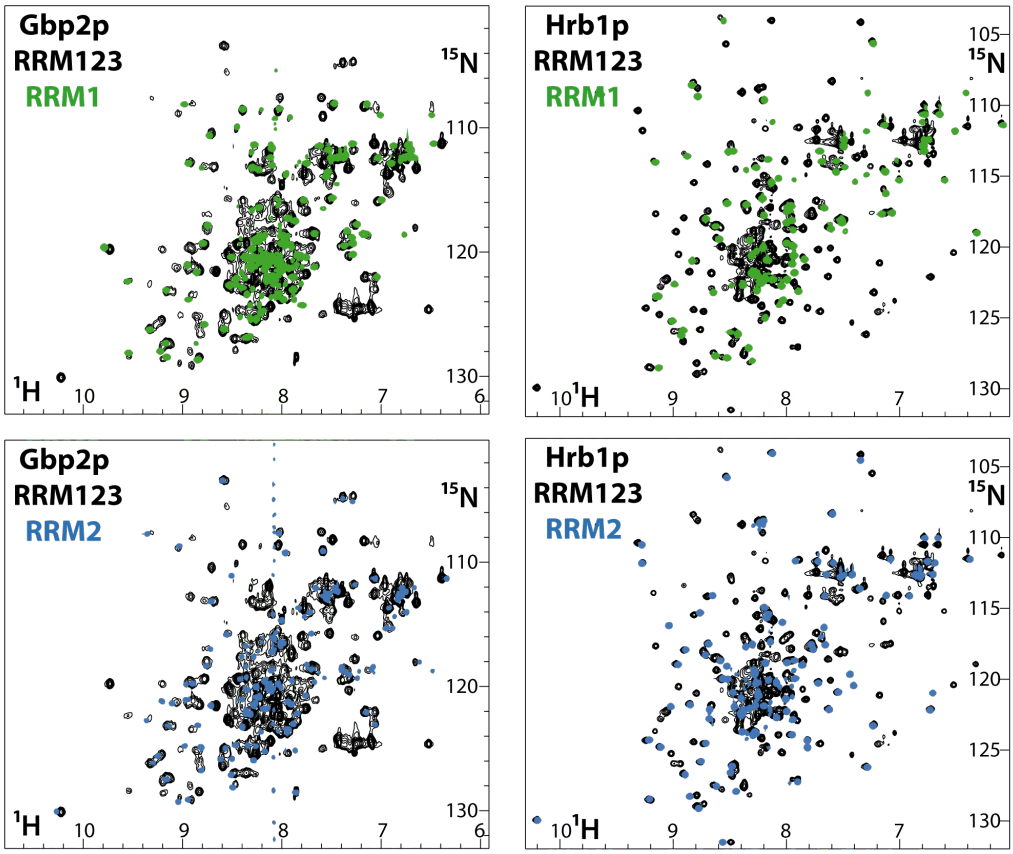


Figure 3.17. – ^{1}H-^{15}N HSQC spectra superposition of Gbp2p and Hrb1p RRM12 tandems (black) with corresponding RRM1 (green spectra in upper panels) and RRM2 domains (blue spectra in lower panels).

To summarize, several important conclusions can be drawn from these compendium of physical chemical and bioinformatics analyses of Hrb1p and Gbp2p proteins. First, the six RRM domains are folded under the experimental conditions presumably in the typical α/β arrangement, but their stabilities are diverse and suggest small structural differences to this canonical fold. The Gbp2p and Hrb1p domain architecture seem constituted by two independent modules: RRM12 tandem and RRM3; where Gbp2p/Hrb1p RRM1 and RRM2 domains build up tandems, which are more stable than the sum of their parts. It is possible that this increased stability is due to the existence of domain-domain interactions (intra or intermolecular) that generates heterogeneity in the NMR data. Gbp2p/Hrb1p RRM3 constructs are the most stable constructs and render high quality NMR data. Despite these two RRM3 constructs contain extra regions outside the canonical definitions of the RRM, their ^{1}H-^{15}N-HSQC spectra do not indicate the presence of large unstructured regions. All these circumstances point out to RRM3 domains as attractive candidates for structural studies.

3.2.4. High-resolution structural studies of Gbp2p and Hrb1p

Previous physical chemical characterization yields interesting information about the internal organization of Gbp2p/Hrb1p proteins. The presence of well defined RRM domains together with evidences of expanded structural motifs (in RRM3) and interdomain interactions, make the determination of the high-resolution structure of both proteins of great interest. Moreover, the availability of high-resolution structures is a pre-requisite to interpret the binding data in terms of definition of molecular interfaces. NMR is a powerful technique to obtain structural information at different levels. The secondary structure composition of Gbp2p and Hrb1p RRM domains was investigated first by using the chemical shift index data (CSI) obtained from protein assignment. The domain boundaries were confirmed in the large constructs by this method. In favourable cases, a full three-dimensional structure determination was performed following the well-established NOE-based NMR methodology for constructs containing only one RRM, for which size and spectral quality was optimal. For longer structural units, like the RRM12 tandem, alternative techniques such as X-ray crystallography were attempted. Finally, inter domain contact analysis was performed for both proteins by NMR spectra comparison between longer constructs (RRM123 and RRM12) and single RRMs.

3.2.4.1. Protein assignment and secondary structure analysis

The NMR chemical shift encodes all the information about the structural environment of a given nucleus and can be used in an empirical manner to generate restraints for structure calculations. In amino acids, backbone nuclei chemical shifts depend on the type of residue (different chemical composition) but it is also strongly correlated with the conformation adopted by the peptide bond (*e.g.* α-helix, β-sheet). Typically, $^{1}H\alpha$, ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts deviate to upfield or downfield values relative to random coil, depending on the secondary structure. The chemical shift index (CSI) is a normalized value (1, 0 or -1) obtained from the differences between backbone chemical shifts and their corresponding random coil reference values (Wishart and Sykes, 1994a, b). A consecutive distribution of CSI of the same sign in the sequence can be interpreted as evidence of α-helix (CSI = -1) or β-sheet (CSI = 1) secondary structure elements. Regions with changes in the CSI or zero values are interpreted as non-regular secondary structure sections (sometimes referred as "coil").

In order to perform the CSI analysis, the backbone chemical shifts were determined by assignment of a set of three-dimensional experiments obtained from different protein constructs of Gbp2p and Hrb1p. Previous NMR data suggested that isolated RRM3 and RRM12 tandem construct accurately represent the structure of the RRM123 one (superposition of RRM123 ^{1}H-^{15}N HSQC spectra with RRM12 and RRM3 ones overlapped almost perfectly). The enormous challenge of getting the RRM123 assignments was simplified by dividing them into two more tractable constructs: RRM12 and RRM3. In the case of Gbp2p, RRM12 construct was not used due to serious line broadening effect (described in 3.2.3.4 section) that compromises NMR spectra quality. The Gbp2p RRM123 construct was studied instead, as it presented an attenuated line broadening effect. In addition, the assignment of RRM1 and RRM2 individual constructs of both proteins was also obtained and in most cases was useful to translate the assignment directly to the longer constructs.

Three-dimensional NMR datasets were acquired using purified ^{13}C and ^{15}N labelled samples of all those constructs (Table 3.2). The concentration of each sample was optimized to maximize signal to noise while avoiding disturbing effects (line broadening, aggregation, etc.) and all samples were prepared in the same buffer (150-800 μM protein in a 25 mM potassium phosphate pH 6.5, 25 mM NaCl and 0.1 mM DTT). A second set of natural abundance protein samples at the same experimental conditions was used in order to obtain another collection of

experiments (Table 3.2). Not all experiments could be obtained for larger constructs due to their serious broadening effect.

	Protein construct	Gbp2p					Hrb1p				
		RRM123	RRM12	RRM1	RRM2	RRM3	RRM123	RRM12	RRM1	RRM2	RRM3
	Concentration (μM)	350	150	200	200	800	450	400	200	200	800
2D	^{15}N HSQC (H_2O / D_2O)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	^{13}C HSQC (D_2O)	✓		✓	✓	✓		✓	✓	✓	✓
2D	CB(CGCD)HD (D_2O)					✓					✓
	CB(CGCGCE)HE (D_2O)					✓					✓
3D	HNCA (H_2O)	✓		✓	✓	✓		✓	✓	✓	✓
	HNCO (H_2O)	✓		✓	✓	✓		✓	✓	✓	✓
	CBCA(CO)NH (H_2O)	✓		✓	✓	✓		✓	✓	✓	✓
	CBCANH (H_2O)					✓					✓
	HN(CA)CO (H_2O)					✓					✓
3D	HC(C)H-TOCSY (D_2O)			✓	✓	✓			✓	✓	✓
	(H)CCH-TOCSY (D_2O)			✓	✓	✓			✓	✓	✓
2D	TOCSY (H_2O)			✓	✓	✓			✓	✓	✓
	TOCSY (D_2O)					✓			✓	✓	✓
	NOESY (H_2O)			✓	✓	✓			✓	✓	✓
	NOESY (D_2O)					✓			✓	✓	✓

Table 3.2. – Different NMR experiments acquired for each construct indicating the concentration and the solvent used in each case (H_2O solvent always contains a 10% of D_2O).

The assignment protocol is described in detail in 2.12.2 section and basically consisted on the connection of backbone nuclei of each residue using the triple-resonance data and then, the completion of the side chain assignment using the 3D HCCH-TOCSY (only for single RRM constructs). RRM2 and RRM3 constructs were almost completely assigned (Table 3.3). The RRM1 domains were assigned at lower percentages because some amino acids belonging to the same region were not observed in both proteins. The extent of backbone assignments (^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα and ^{13}Cβ) for the largest constructs was also high. The assignment of Hrb1p RRM12 and Gbp2p RRM123 was made by a combination of triple resonance data and comparison with the assignment of the individual domains, while Hrb1p RRM123 and Gbp2p RRM12 ^{1}H-^{15}N HSQC spectra were assigned exclusively by comparison.

Construct		Backbone nuclei					Whole assignment		
		$^{15}N_H$	$^{1}H_N$	^{13}CO	$^{13}C\alpha$	$^{13}C\beta$	^{15}N	^{1}H	^{13}C
Gbp2p	**RRM1**	72.5	73.3	73.6	75.8	81.2	54.6	64.9	63.6
	RRM2	98.0	100.0	97.0	100.0	100.0	70.5	78.2	75.3
	RRM3	92.9	97.9	100.0	100.0	100.0	86.4	95.4	86.4
	RRM12	76.4	78.6	-	-	-	-	-	-
	RRM123	83.5	86.7	57.6	73.2	50.3	-	-	-
Hrb1p	**RRM1**	86.0	88.0	76.3	94.6	93.1	64.7	83.8	79.3
	RRM2	94.8	98.9	92.8	100.0	100.0	82.3	94.2	88.7
	RRM3	98.0	100.0	100.0	100.0	100.0	87.8	96.1	90.8
	RRM12	92.3	95.1	-	-	-	-	-	-
	RRM123	90.2	93.3	77.2	95.8	87.5	-	-	-

Table 3.3. – Percentage of assignment completeness for each construct, differencing the backbone nuclei and the global nuclei assignment.

Once backbone nuclei were assigned, consensus chemical shift index (CSI) was calculated using all $^{1}H\alpha$ (except for Gbp2p RRM123 and Hrb1p RRM12), ^{13}CO, $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts for each construct and the results are shown in Figure 3.18.

In agreement with the bioinformatics analysis, RRM1 and RRM2 of both proteins exhibit the typical RRM fold ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$); the last β-strand is ill-defined in some cases, probably due to its shortness. In contrast, both RRM3 domains contain an additional α-helix in the N-terminal region that precedes the regular RRM domain; this was also predicted in the bioinformatics analysis. Finally, the obtained domain boundaries match with those proposed in databases and no other regions seem to show any regular secondary structure. Moreover, peak intensities and signal dispersion of residues in interdomain linkers and in the N-terminal region of RRM1 domains are characteristic of intrinsically unstructured proteins.

It could be clearly observed that the RRM arrangement is conserved along all constructs (in large ones and in the single RRM constructs), thus to simplify the study, a structural calculation of the smallest constructs was carried out. The quality of the data varied between the protein constructs; RRM3 domains present excellent sets of spectra facilitating the structure calculation whereas RRM1 and RRM2 spectra show line broadening effects, being critical in the case of Gbp2p, that complicates the structure determination of these domains.

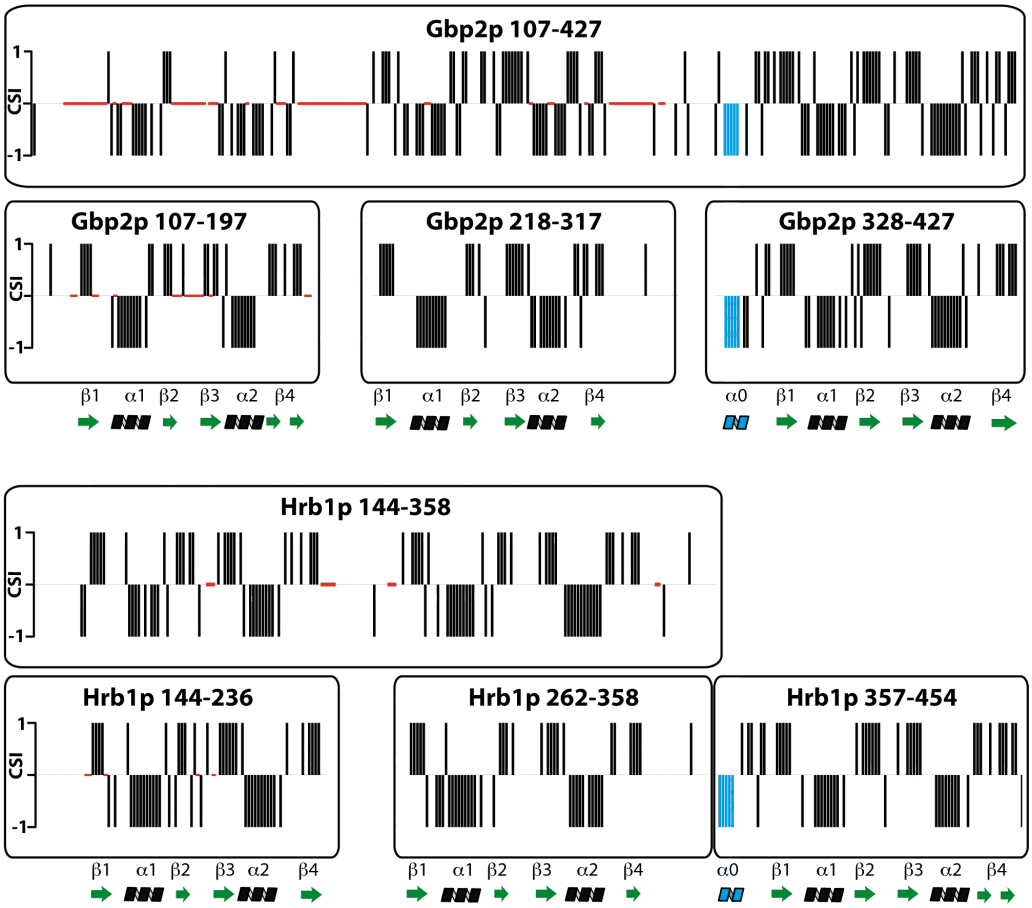


Figure 3.18. – Consensus CSI for different constructs of Gbp2p and Hrb1p. Values of -1 point to α-helix prediction, +1 to β-strands and 0 values indicate non secondary structure preference. At least four consecutive residues with CSI = -1 have been required for α-helix prediction, three or more with CSI = +1 in the case of β-strands. Red marks indicate unassigned amino acid positions and blue bars the non canonical RRM elements.

3.2.4.2. Structural study of single RRM constructs by Nuclear Magnetic Resonance

The three-dimensional structures of Hrb1p RRM1, RRM2, RRM3 and Gbp2p RRM3 constructs were calculated from NMR data; Gbp2p RRM1 and RRM2 structures were not obtained due to their bad spectra quality, as it was previously pointed out. Restraint lists (Table 3.4) were prepared using information of NOE-derived distance (obtained from 2D NOESY spectra) and angular (generated by TALOS+) restraints. Lists of both types of constraints were implemented in CYANA 2.1 program to calculate an ensemble of twenty structures for each domain and a final molecular dynamics step was performed using the Amber 9.0 force field

(see 2.12.3 section for details). Table 3.4 summarises the structure calculation statistics and shows the quality differences between RRM3 domains and RRM1/RRM2 structures. The 2D NOESY experiments of RRM1 and RRM2 domains were acquired at low protein concentration, in order to minimize line broadening effects, thus their signal to noise, and therefore the number of NOE-derived distance restrictions used, is substantially lower than the RRM3 equivalents. In addition, in the spectra of Hrb1p RRM1 domain a large histidine-rich segment shows severe line broadening, this time probably due to coupled acid-base equilibrium among the histidines (as it happens with histidine-tags in fusion proteins). Because of this, the entire loop 3 was unassigned and is mostly undefined (Figura 3.19). Excluding this region from the comparison, the RMSD values drop to values more comparable with the other structures (Table 3.5), showing that the domain core remains well structured.

As expected, the calculated structures (Figure 3.19) display the $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ global fold, characterized by the packing of the two α-helices against one side of the antiparallel β-sheet. Both RRM3 domains (only Hrb1p RRM3 is shown in the figure) contain a non-canonical extra α-helix at the N-terminus, packed against the other β-sheet face. The segment connecting this novel element to the RRM fold is wrapped around the C-terminal end of the protein in a closed loop conformation.

The folded region of Hrb1p RRM1 construct domain spans from residues Ser 162 to Asn 236 (at the end of the construct), whereas Hrb1p RRM2 is structured from Glu 262 (at the beginning of the construct) to Lys 336. The N-terminal residues of RRM1 construct (Ala 144–Asn 161) and the C-terminal part of the RRM2 domain (Glu 337- Asn 358) seem to be disordered in consistency with the lack of medium/long-range NOESY peaks in these regions, the narrow and intense ^{1}H-^{15}N-HSQC peaks and the poor chemical shift dispersion of their signals (as in unstructured peptides). In contrasts, RRM3 constructs of both proteins are completely structured, except the first pair of residues.

	Hrb1p RRM1	Hrb1p RRM2	Hrb1p RRM3	Gbp2p RRM3
NMR experimental restraints				
NOE-derived				
Intraresidue	274	190	589	533
Sequential	163	137	540	521
Medium-range (1 < i - j < 4)	164	131	476	476
Long-range (i - j > 4)	375	283	1312	1370
Total per residue	13.0	10.0	30.1	30.2
TALOS+ obtained restraints				
ϕ angle restraints	49	58	70	71
ψ angle restraints	51	56	69	68
Structure statistics				
Mean AMBER energies (Kcal/mol ± SD)				
Total	-3772 ± 23	-2537 ± 12	-3300 ± 11	-2925 ± 12
Van der Waals	-643 ± 15	-599 ± 10	-764 ± 6	-751 ± 7
Restrains (distance + angle)	4 ± 1	6 ± 1	27 ± 1	19 ± 1
Violations				
Distance*	0.2 ± 0.4	1.1 ± 0.2	1.5 ± 1.0	1.4 ± 0.9
Maximum distance violation (Å)	0.15	0.20	0.26	0.17
Angle**	0.5 ± 0.7	0.5 ± 0.7	0.8 ± 0.9	0.2 ± 0.4
Maximum angle violation (°)	7.6	7.4	6.6	3.6
RMSD from ideal geometry				
Bond lengths (Å)	0.010	0.010	0.009	0.009
Bond angles (°)	1.98	2.22	1.85	2.05
Averages RMSD to mean structure (range)	***(162-236)***	***(262 - 336)***	***(357-454)***	***(330-427)***
N, CO, Cα (Å) (± SD)	1.39 ± 0.40	0.72 ± 0.19	0.33 ± 0.09	0.43 ± 0.08
All heavy (Å) (± SD)	4.47 ± 2.18	3.87 ± 1.63	0.72 ± 0.09	0.85 ± 0.09

Table 3.4. – Summary of NMR restraint and structural calculation statistics for different domain structures of Gbp2p and Hrb1p. * Averaged value per structure of distance violations > 0.15 Å ± SD. ** Averaged value per structure of total angle violations ± SD.

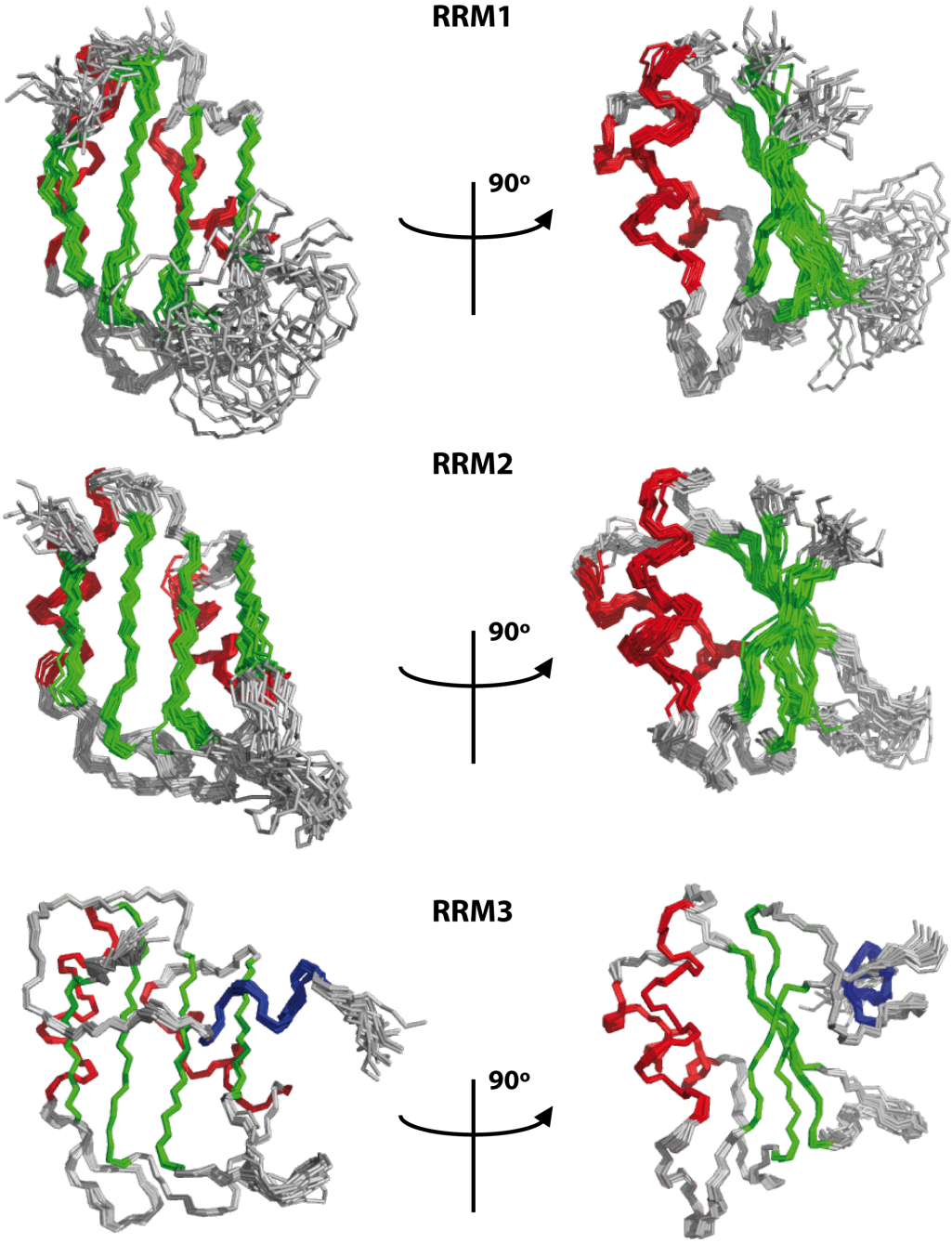


Figure 3.19. – Superposition of the 20 calculated structures for Hrb1p RRM1 (upper panel), Hrb1p RRM2 (middle panel) and Hrb1p RRM3 (lower panel). Secondary structure are coloured in green (β-sheet), red (α-helix) and blue (α_0 helix in Hrb1p RRM3). PDB files of the structures (these and Gbp2p RRM3) are distributed in supplementary material (CD).

Considering only the regular secondary structure elements, RMSD values for the NMR ensembles of single RRM domain constructs are below 0.9 Å (< 0.3 Å for the two RRM3), indicating that the structures are well defined (Table 3.5 and Figure 3.19). Excluding the loops, particularly the disordered loop 3 (due to the lack of restraint), RMSD value of RRM1 decreases significantly (compared with the value reported in Table 3.4). The RMSD values between pairs of different RRM domains are not so low (Table 3.5 and Figure 3.20), except for the two RRM3 domains with an RMSD value around 0.9 Å.

	Hrb1p RRM1	Hrb1p RRM2	Hrb1p RRM3	Gbp2p RRM3
Hrb1p RRM1	**0.90 ± 0.17**	**1.64 ± 0.14**	**1.49 ± 0.06**	**1.55 ± 0.84**
Hrb1p RRM2		**0.85 ± 0.14**	**1.39 ± 0.09**	**1.42 ± 0.10**
Hrb1p RRM3			**0.19 ± 0.04**	**0.91 ± 0.04**
Gbp2p RRM3				**0.29 ± 0.07**

Table 3.5. - RMSD (in Å) obtained from the pair wise alignment between structures within NMR ensembles of Hrb1p RRM1, RRM2, RRM3 and Gbp2p RRM3. Backbone atoms from the α-helices and β-sheet were used in the alignment. The following residue ranges were used: Hrb1p RRM1 (β_1:161-167, β_2:189-192, β_3:201-206, β_4:231-234, α_1:175-181 and α_2:211-221), Hrb1p RRM2 (β_1:261-267, β_2:289-292, β_3:302-307, β_4:332-335, α_1:275-281 and α_2:312-322), Hrb1p RRM3 (β_1:376-382, β_2:404-407, β_3:417-422, β_4:447-450, α_1:390-396 and α_2:427-437) and Gbp2p RRM3 (β_1:349-355, β_2:377-380, β_3:390-395, β_4:420-423, α_1:363-369 and α_2:400-410).

In fact, the structures of the two RRM3 are very similar displaying only slight differences on the conformation of loop 3 (Figure 3.20 upper panel). This loop is solvent exposed, without contacts with other regions and its sequence differs between Hrb1p and Gbp2p, which are otherwise very similar (69% of identity, see Appendix 2). The global fold of Hrb1p RRM1 resembles that of Hrb1p RRM3, including the conformation of all loops, except for the undefined loop 3 (see above) and, obviously, excluding the extra N-terminal region. In contrast, Hrb1p RRM2 differs from these two in the conformation of loop 2, which is in an open-to-the-solvent conformation in this case. In general, the central elements comprising the RRM fold are similarly oriented in the four structures (Table 3.5 and Figure 3.20).

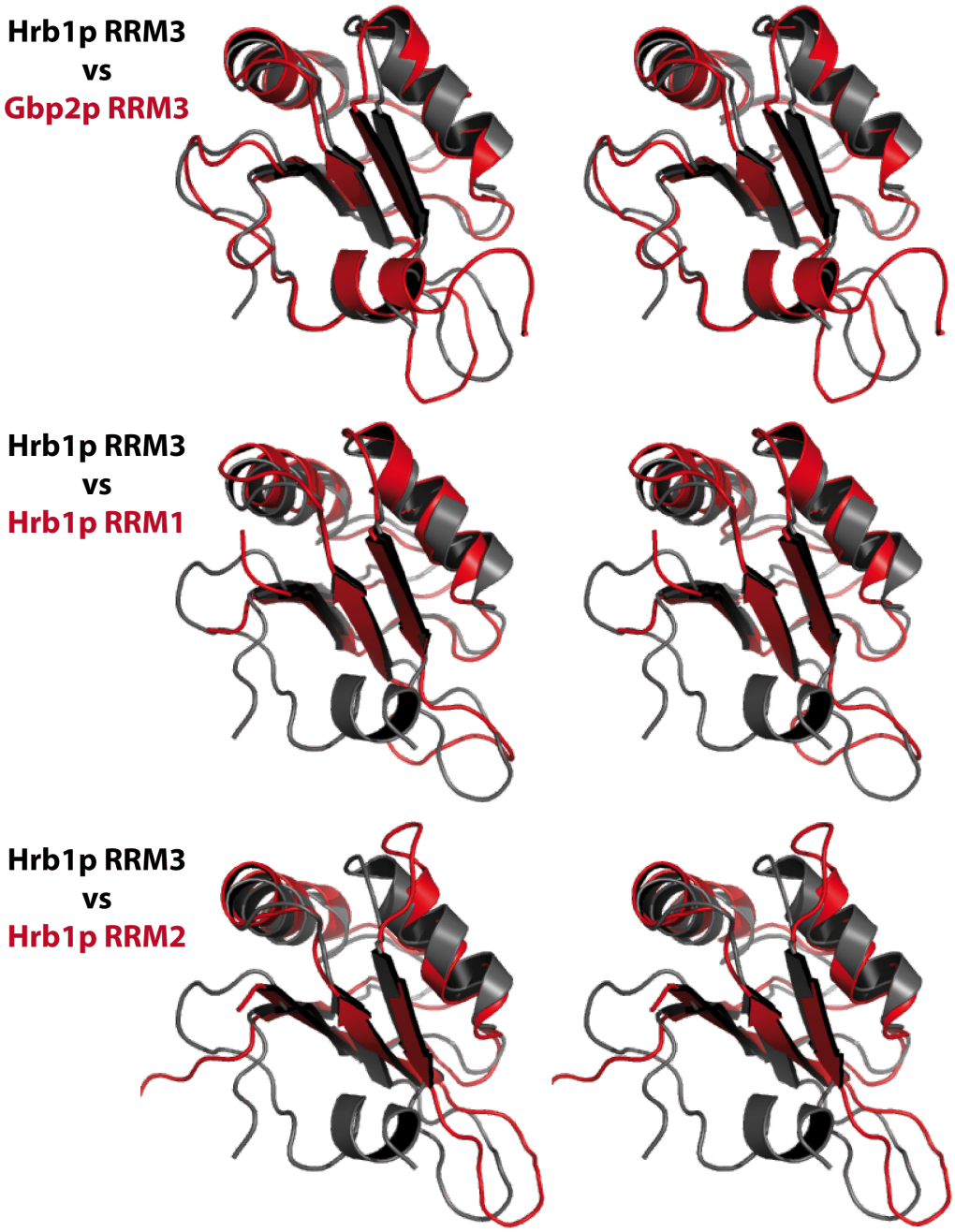


Figure 3.20. - Cross-eyed stereo views of the structural alignments between Gbp2p RRM3 (in red in upper panel), Hrb1p RRM2 (in red in middle panel) and Hrb1p RRM1 (in red in lower panel), and Hrb1p RRM3 (in black in all panels). Regions used for fitting are described in Table 3.5.

3.2.4.2.1. Secondary structure and RRM fold definition

Figure 3.21 represents a scheme of the four-stranded β-sheet in RRM domains, introducing a new general numbering code that refers to different locus on the β-sheet, which is useful to compare different aspects of RRM architecture and stability among the structures. This code has been chosen to define even loci to residue positions whose side chains are pointing toward the buried face of the β-sheet (odd positions side chains are solvent exposed and even ones buried). The loci 33 and 35 in RNP1 and 13 in RNP2 identify the positions typically involved in RNA binding (Clery et al., 2008; Daubner et al., 2013) (Figure 1.5 and 1.6) which are solvent exposed.

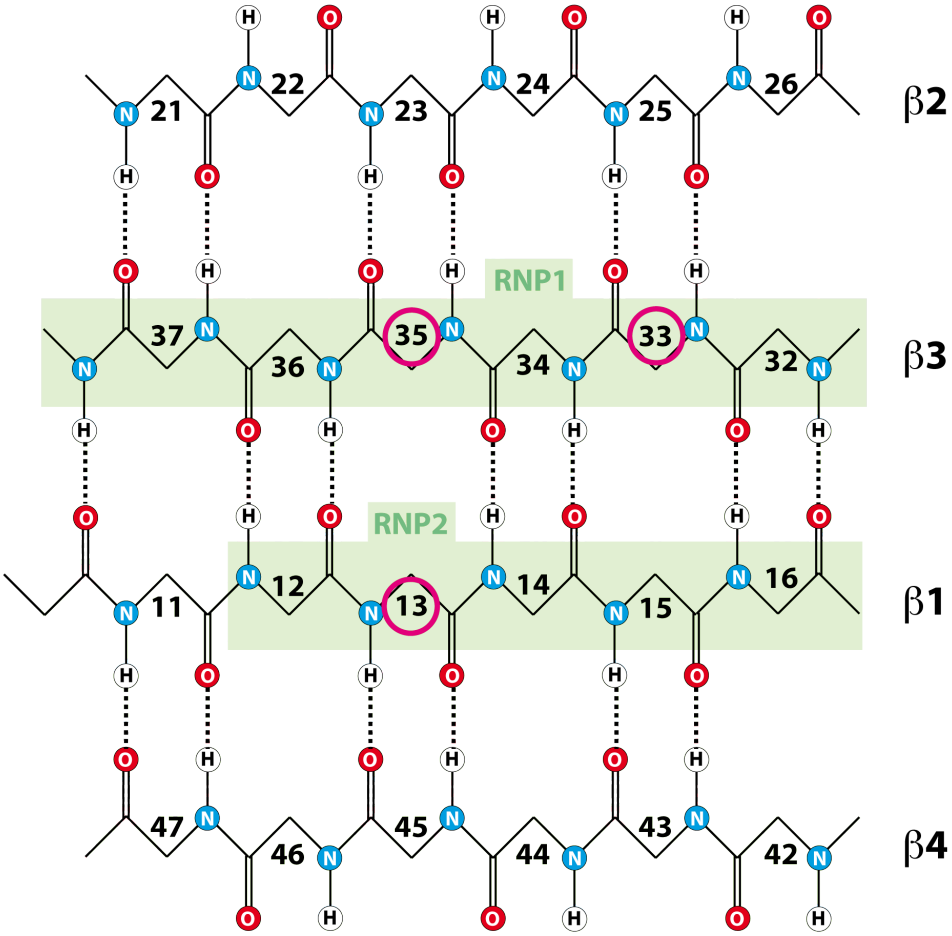


Figure 3.21. - Schematic representation of the RRM β-sheet. The characteristics RNP1 and RNP2 motifs have been shadowed. Numbers corresponding to usual RNA-binding loci have been circled in magenta.

A network of hydrogen bonds stabilizes the central β-sheet of the RRMs. The stability of a given RRM is correlated with the solvent exchange rates of the amide protons involved in this hydrogen bond network. Protected amide protons can be easily identified by comparing the spectra recorded in H_2O or in D_2O solvents (Figure 3.22).

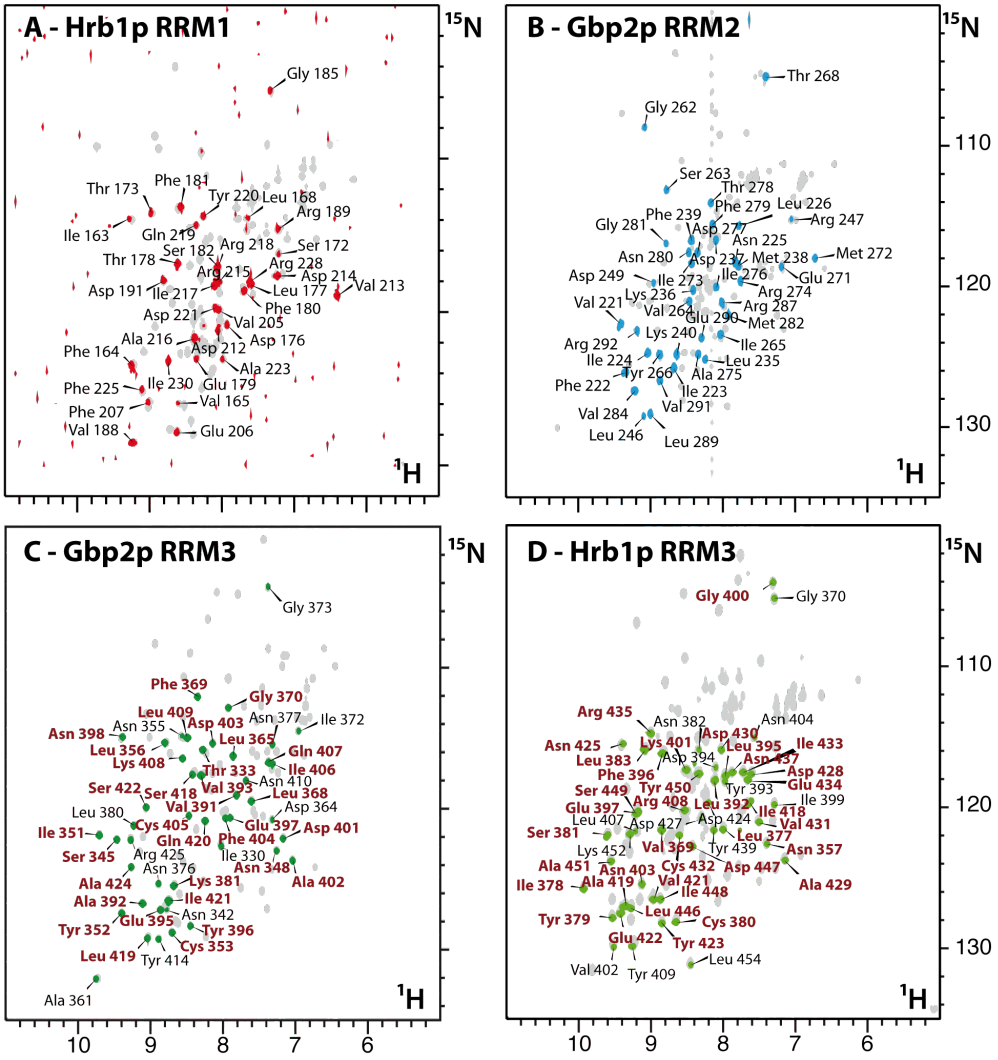


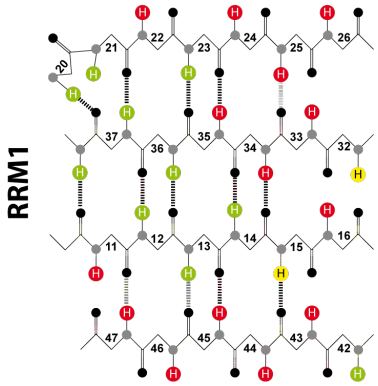
Figure 3.22. – Solvent exchange data for various Hrb1p/Gbp2p RRM domains monitored on the ^{1}H-^{15}N HSQC spectra (A – Hrb1p RRM1, B – Gbp2p RRM2, C – Gbp2p RRM3 and D - Hrb1p RRM3). The experiments have been recorded after ~ 1h of dissolving the lyophilized sample in D_2O; peaks of residues with solvent protected amide groups are labelled and those remaining after 1 day of exchange are highlighted in brown bold letters. The reference spectra in H_2O are shown in grey on the background of each panel.

The solvent exchange data obtained for the different Gbp2p/Hrb1p constructs present important differences. Up to 34 backbone amide crosspeaks peaks can be observed in the ^{1}H-^{15}N HSQC spectrum of Hrb1p RRM1 recorded < 1h after dissolving the lyophilized sample in 100% D_2O (43% of the HN assigned in H_2O) (Figure 3.22.A). In contrast, all amides of Gbp2p RRM1 exchange within the dead time of the experiment (< 1h after solving in D_2O). The opposite scenario happens for RRM2 domains; all Hrb1p RRM2 amides exchange very fast and 39 are protected in Gbp2p RRM2 (41% of the HN assigned in H_2O) (Figure 3.22.B). The two RRM3 domains are clearly the most stable among the RRM domains (as already seen by thermal denaturation; see 3.2.3.3 section), because a larger percentage of crosspeaks survive in D_2O (Figure 3.22.C-D): 51 backbone amides in the case of Hrb1p RRM3 (53% of the total) and 46 for Gbp2p RRM3 (51% of the total). The lifespan is longer in RRM3 domains; 38 and 34 crosspeaks remain in Hrb1p/Gbp2p RRM3 after one day in D_2O; a period in which all Gbp2p RRM2 and Hrb1p RRM1 crosspeaks have disappeared from the spectrum.

In Gbp2p/Hrb1p RRM3 constructs, all the backbone amides involved in hydrogen bonds in the β-sheet are protected. In contrast, the hydrogen bond positions in the β-sheet of Gbp2p RRM2 and Hrb1p RRM1 are more accessible to the solvent in particular those that are closer to the edges (Figure 3.23). Moreover, none of the amide protons in hydrogen bond positions in the fourth strand of Hrb1p RRM1 are protected. The β-sheets in the four calculated structures also present the canonical geometry (N-H-O angles between 0 and 35º and H-O distance up to 2.4 Å, detected by molmol software) in most of the hydrogen bonds positions (Figure 3.23).

Curiously, a β-bulge was observed in β_2 of Hrb1p RRM1 domain. In the N-terminal of the strand, Val 188 (locus 20) places its H_N inside the interstrand region instead of pointing out of the β-sheet, disrupting the regular hydrogen bonding: Arg 189 H_N (locus 21) does not present the expected interaction with Glu 206 CO (locus 37), which forms a hydrogen bond with Val 188 H_N (present in all NMR calculated structures and also supported by different NOESY crosspeaks).

Apart form the H_N of the β-sheets, other amide protons are protected, mainly on the α-helixes (Figure 3.24). The two helixes of the RRM fold are well defined in the four calculated structures and their lengths are around 9 residues in α_1 and 11 in α_2. They are mostly in the α-helix conformation, characterized by the hydrogen bond pattern CO(i)-H_N(i+4).



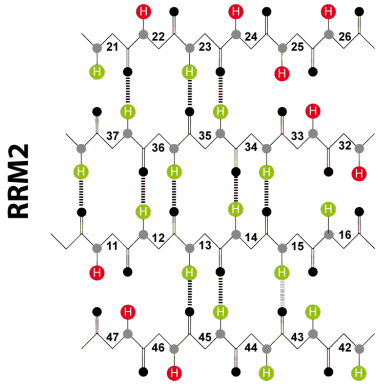
RRM1

		21	22	23	24	25	26
β2	Gbp2p	Glu 150	Ala 151	Asp 152	Ile 153	Ile 154	Thr 155
	Hrb1p	Arg 189	Ala 190	Asp 191	Ile 192	Ile 193	Thr 194

		37	36	35	34	33	32
β3	Gbp2p	Glu 167	Val 166	Thr 165	Gly 164	Met 163	Gly 162
	Hrb1p	Glu 206	Val 205	Thr 204	Gly 203	Met 202	Gly 201

		11	12	13	14	15	16
β1	Gbp2p	Ser 123	Ile 124	Phe 125	Val 126	Arg 127	Asn 128
	Hrb1p	Ser 162	Ile 163	Phe 164	Val 165	Gly 166	Asn 167

		47	46	45	44	43	42
β4	Gbp2p	Asp 196	Gln 195	Arg 194	Val 193	Met 192	Leu 191
	Hrb1p	Asp 235	Gln 234	Arg 233	Val 232	Phe 231	Ile 230



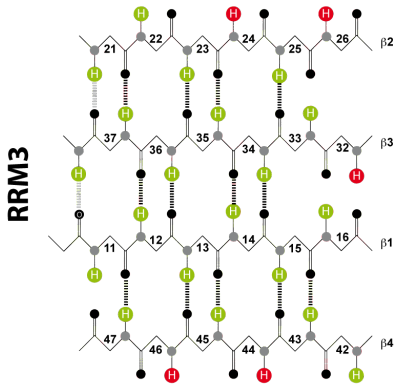
RRM2

		21	22	23	24	25	26
β2	Gbp2p	Arg 247	Ala 248	Asp 249	Val 250	Glu 251	Leu 252
	Hrb1p	His 289	Ala 290	Asp 291	Val 292	Glu 293	Leu 294

		37	36	35	34	33	32
β3	Gbp2p	Ile 265	Val 264	Ser 263	Gly 262	Phe 261	Gly 260
	Hrb1p	Ser 307	Val 306	Thr 305	Gly 304	Ser 303	Gly 302

		11	12	13	14	15	16
β1	Gbp2p	Glu 220	Val 221	Phe 222	Ile 223	Ile 224	Asn 225
	Hrb1p	Glu 262	Val 263	Ile 264	Val 265	Lys 266	Asn 267

		47	46	45	44	43	42
β4	Gbp2p	Gly 294	Glu 293	Arg 292	Val 291	Glu 290	Leu 289
	Hrb1p	Lys 336	Ser 335	Lys 334	Val 333	Asp 332	Leu 331



RRM3

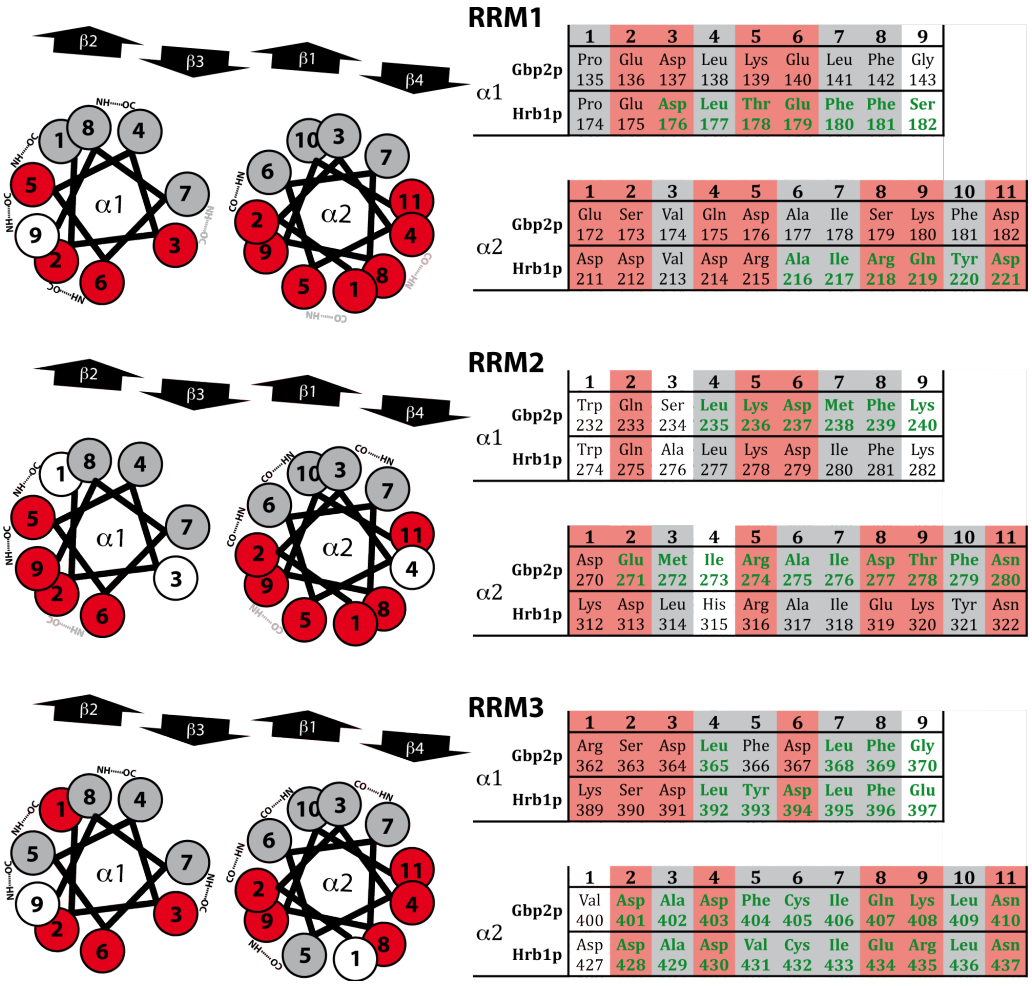
		21	22	23	24	25	26
β2	Gbp2p	Asn 377	Ala 378	Glu 379	Leu 380	Lys 381	Pro 382
	Hrb1p	Asn 404	Ala 405	Glu 406	Leu 407	Arg 408	Tyr 409

		37	36	35	34	33	32
β3	Gbp2p	Glu 395	Val 394	Val 393	Ala 392	Val 391	Gly 390
	Hrb1p	Glu 422	Val 421	Val 420	Ala 419	Ile 418	Gly 417

		11	12	13	14	15	16
β1	Gbp2p	Phe 350	Ile 351	Tyr 352	Cys 353	Ser 354	Asn 355
	Hrb1p	Leu 377	Ile 378	Tyr 379	Cys 380	Ser 381	Asn 382

		47	46	45	44	43	42
β4	Gbp2p	Ala 424	Tyr 423	Ser 422	Ile 421	Gln 420	Leu 419
	Hrb1p	Ala 451	Tyr 450	Ser 449	Ile 448	Asp 447	Leu 446

Figure 3.23. - Schematic representation of the H/D exchange data for the amide protons involved in hydrogen bonds in the β-sheet of all RRMs. Exchangeable protons are in red and protected ones in green. Canonical hydrogen bond geometries detected by MOLMOL are drawn for each pair (black for those detected in more than 15 structures and grey in 10-15). Up panel for RRM1 (data from Hrb1p), middle for RRM2 (H/D data from Gbp2p and structures from Hrb1p) and down for RRM3 domain (combined data from both proteins).



RRM1

		1	2	3	4	5	6	7	8	9
α1	Gbp2p	Pro 135	Glu 136	Asp 137	Leu 138	Lys 139	Glu 140	Leu 141	Phe 142	Gly 143
	Hrb1p	Pro 174	Glu 175	**Asp 176**	**Leu 177**	**Thr 178**	**Glu 179**	**Phe 180**	**Phe 181**	**Ser 182**

		1	2	3	4	5	6	7	8	9	10	11
α2	Gbp2p	Glu 172	Ser 173	Val 174	Gln 175	Asp 176	Ala 177	Ile 178	Ser 179	Lys 180	Phe 181	Asp 182
	Hrb1p	Asp 211	Asp 212	Val 213	Asp 214	Arg 215	**Ala 216**	**Ile 217**	**Arg 218**	**Gln 219**	**Tyr 220**	**Asp 221**

RRM2

		1	2	3	4	5	6	7	8	9
α1	Gbp2p	Trp 232	Gln 233	Ser 234	**Leu 235**	**Lys 236**	**Asp 237**	**Met 238**	**Phe 239**	**Lys 240**
	Hrb1p	Trp 274	Gln 275	Ala 276	Leu 277	Lys 278	Asp 279	Ile 280	Phe 281	Lys 282

		1	2	3	4	5	6	7	8	9	10	11
α2	Gbp2p	Asp 270	**Glu 271**	**Met 272**	**Ile 273**	**Arg 274**	**Ala 275**	**Ile 276**	**Asp 277**	**Thr 278**	**Phe 279**	**Asn 280**
	Hrb1p	Lys 312	Asp 313	Leu 314	His 315	Arg 316	Ala 317	Ile 318	Glu 319	Lys 320	Tyr 321	Asn 322

RRM3

		1	2	3	4	5	6	7	8	9
α1	Gbp2p	Arg 362	Ser 363	Asp 364	**Leu 365**	Phe 366	Asp 367	**Leu 368**	**Phe 369**	**Gly 370**
	Hrb1p	Lys 389	Ser 390	Asp 391	**Leu 392**	**Tyr 393**	**Asp 394**	**Leu 395**	**Phe 396**	**Glu 397**

		1	2	3	4	5	6	7	8	9	10	11
α2	Gbp2p	Val 400	**Asp 401**	**Ala 402**	**Asp 403**	**Phe 404**	**Cys 405**	**Ile 406**	**Gln 407**	**Lys 408**	**Leu 409**	**Asn 410**
	Hrb1p	Asp 427	**Asp 428**	**Ala 429**	**Asp 430**	**Val 431**	**Cys 432**	**Ile 433**	**Glu 434**	**Arg 435**	**Leu 436**	**Asn 437**

Figure 3.24. - Schematic representation of the canonical α-helixes in RRM domains (RRM1 up, RRM2 middle and RRM3 down). Hydrophobic residues are drawn in grey and charged and polar ones in red. Residues with non-exchangeable (H/D) amide protons are written in green along the tables, and hydrogen bonds predicted by MOLMOL are pointed in the figures (detected in >15 structures in black, between 10 and 15 structures in grey).

All canonical α-helixes present an amphipathic residue distribution, where the positions 4, 7 and 8 of α_1 and 3, 6, 7 and 10 of α_2 contain apolar amino acids and the rest are mostly occupied by polar residues. Both RRM3 α_2 helixes have an exposed hydrophobic residue (in position 5) and RRM2 domain of Gbp2p has a solvent exposed isoleucine (Ile 273) in position 3 of α_2-helix. A distinct feature of the two RRM2 domains is the presence of a tryptophan residue in position 1 of α_1-helix, whose side chain is solvent exposed (as evidenced by fluorescent experiments) and placed next to the edge of the β-sheet (near β_2-strand).

The cores of the four structures maintain the hydrophobic character for most buried residues in the β-sheet (with even numbers of loci in Figure 3.23) and in the α-helixes (grey residues next to the β-sheet in Figure 3.24). Although in some cases, residues with a polar group are buried into the hydrophobic core, as it happens in β-sheet positions 16, 26 and 46 of some RRMs, places next to an edge of the sheet. This situation might contribute to decrease the stability of the RRMs in some cases.

Interestingly, RRM3 domains contain two conserved cysteines buried in the hydrophobic core: Cys 353 (Gbp2p) and Cys 380 (Hrb1p) (at the locus 14 of Figure 3.23) and Cys 405 (Gbp2p) and Cys 432 (Hrb1p) both in the internal face of α_2-helix (position 6 in Figure 3.24). These cysteines do not form disulphide bridge and their Hγ could be observed in the 2D NOESY spectra (in H_2O) (Table 3.6). None of these sulfhydryl protons show NOEs with the water, suggesting that there are not structural water molecules involved on their stabilization. Instead the side chain of Cys 353 (Cys 380 in Hrb1p) is surrounded by hydrophobic residues, but within hydrogen bond range to Gln 420 O. The other cysteine SH group, Cys 405 in Gbp2p (Cys 432 in Hrb1p), is nearby the phenol group of Tyr 396, but the hydrogen bond is better described in Hrb1p (Cys 432 Hγ-Tyr 324 Oη).

Type of proton	Gbp2p RRM3		Hrb1p RRM3		Protection due to
Arg Hη	Arg 347	*6.82 ppm*	Arg 374	*6.91 ppm*	*Hydrogen bond*
Cys Hγ	Cys 353	*1.43 ppm*	Cys 380	*1.54 ppm*	*Buried*
	Cys 405	*1.83 ppm*	Cys 432	*2.07 ppm*	*Buried*
Ser Hγ	Ser 422	*5.47 ppm*	Ser 449	*6.26 ppm*	*Buried*
Thr Hγ1	Thr 360	*6.13 ppm*	Thr 387	*6.18 ppm*	*Hydrogen bond*
	-		Thr 398	*5.43 ppm*	*Hydrogen bond*
	Thr 389	*5.32 ppm*	Thr 416	*4.78 ppm*	*Hydrogen bond*
Tyr Hη	Tyr 396	*7.75 ppm*	-		*Buried*

Table 3.6. – Non-amide slow exchangeable protons identified in the Gbp2p and Hrb1p RRM3 spectra, and their chemical shift values. Residues in the same line are conserved positions along Gbp2p and Hrb1p sequences.

Loop regions also contain solvent exchange protected N_H groups. The conformation of Hrb1p RRM1 loop 5 as well as those of loops 1,3, and 5 in RRM3 structures are defined by a complex hydrogen bond network (Figure 3.25). In the later case, the network also involve hydroxyl groups that present slow exchange with water (Thr 360 Hγ1 in Gbp2p and Thr 387 Hγ1 in Hrb1p in loop 1 and Thr 389 Hγ1 of Gbp2p and Thr 416 Hγ1 of Hrb1p in loop 3; Table 3.6).

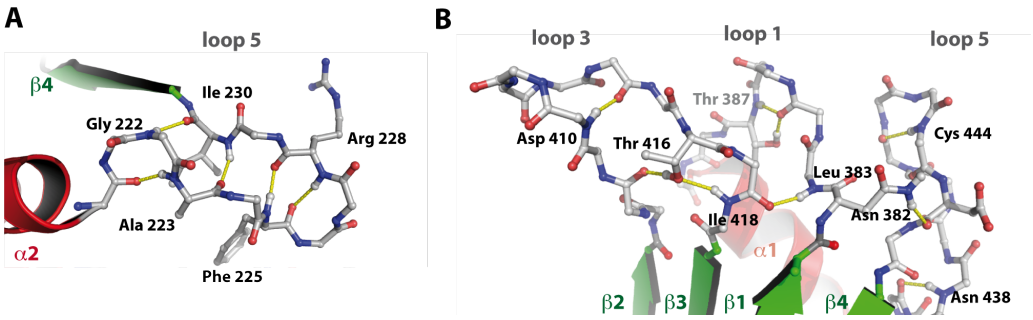


Figure 3.25. – A) Conformation of loop 5 in Hrb1p RRM1 domain. B) Detail of the interaction network in Hrb1p RRM3 loops 1, 3 and 5. Hydrogen bonds detected in more than 15 structures are drawn in yellow and residues with protected N_H labelled in red. Those amino acids whose hydrogen atoms participate in hydrogen bond formation are labelled.

3.2.4.2.2. Novel structural elements on Gbp2p/Hrb1p RRM3 domains

The N-terminal region (including the α_0-helix) of RRM3 domains together with the C-terminal one of the domain (and of the protein) display unique interactions that had never been observed in other RRM. In both proteins, the peptide chains adopt a N-terminal α-helix (residues 360 to 365 in Hrb1p and residues 332 to 338 in Gbp2p) (Figure 3.26.B) followed by a small β-strand (β_0 from Asn 367 to Val 369 in Hrb1p and from Arg 425 to Asp 427 in Gbp2p) that builds up a small two-stranded β-sheet with the C-terminus (β_5) of the protein (Figure 3.26.C). This additional mini β-sheet (better observed in Hrb1p structure) is defined by up to three interstrand hydrogen bonds: Asn 348 NH_2 –Ala 424 CO and Arg 425 NH – Asn 342 CO in Gbp2p; Leu 454 HN – Asn 367 CO, Val 369 NH – Lys 452 CO and Lys 452 HN – Val 369 CO in Hrb1p. Among other spectroscopic evidences, the characteristic interstrand Hα-Hα NOE peaks could be observed between Val 368 and Arg 453 residues in Hrb1p (Figure 3.26.C).

The last part of the N-terminal region preceding the canonical RRM fold, is composed by a glycine rich sequence (in Gbp2p 344 Gly and 345 Gly and in Hrb1p 370 Gly, 371 Gly and 372 Gly), that defines a closed loop around the C-terminal region of the domain, which is stabilized by two hydrogen bonds between two loop glycines and a serine and a threonine from the C-terminus (Figure 3.26.A). The loop also contains a conserved arginine (Arg 347 in Gbp2p and 374 Arg in Hrb1p), whose side chain interacts with the α_2-helix (Leu 399 CO in Gbp2p and Asp 427 Oδ1 in Hrb1p) and with a tyrosine residue of β_4-strand (Tyr 423 in Gbp2p and Tyr 450 in Hrb1p). The arginine Hε is hydrogen bonded to the phenol group of the tyrosine (*e.g.* Arg 374 Hε – Tyr 450 Oη, in Hrb1p) slowing down considerably the ring flipping movement

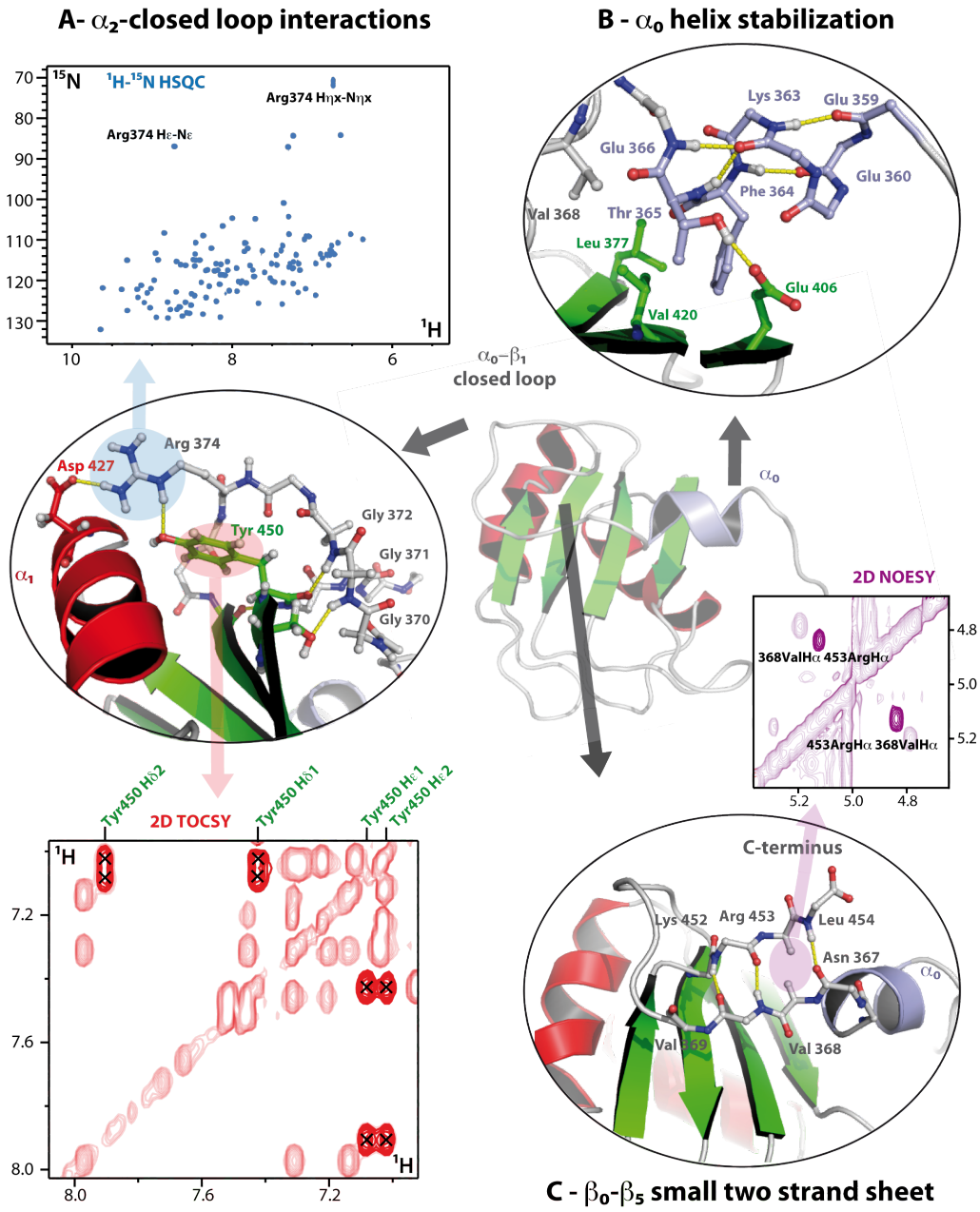


Figure 3.26. – Structural details of the N-terminal novel element in Hrb1p RRM3 domain (equivalent to Gbp2p RRM3). A) The closed loop conformation formed around the C-terminus, indicating the stabilizing interactions directed by the conserved arginine residue (Arg 374); also it is represented the ^{1}H-^{15}N HSQC spectra in order to localize the Hη and the shifted Hε protons of the arginine (up) and a detail of the 2D TOCSY and ^{1}H-^{13}C HSQC spectra showing the four signals of each aromatic proton of fixed Tyr 450 (down). B) The stabilization of the α_0-helix through hydrogen bonds and hydrophobic interactions. C) The mini β-sheet formed with C-terminus showing the characteristic of Hα-Hα noe peak between Val 368 and Arg 453 residues (up).

and allowing the observation of the four ring protons of these residues by NMR (for both Gbp2p and Hrb1p domains) (Figure 3.26.A). This Arg Hε chemical shifts are shifted down-field due to this interaction. The Hηxx guanidinium protons of this arginine were also detected in ^{1}H-^{15}N HSQC spectra, confirming that their interaction with the α_2-helix is strong enough to slow their water exchange rate. The fact that in Hrb1p this arginine is flanked by two aspartate residues indicates that the interaction might be largely electrostatic. In Gbp2, Arg 347 Hη21 is hydrogen bonded to Leu 399 O from α_2-helix.

The N-terminal section of RRM3 is anchored to the β-sheet by a small hydrophobic pocket, where two apolar residues of the α_0-helix (in Gbp2p, Phe 337 and Val 341 and in Hrb1p Phe 364 and Val 368) are surrounded by other hydrophobic residues from the β-sheet (in Gbp2p Phe 350, Val 393 and in Hrb1p Leu 377 and Val 420) and by an hydrogen bond formation between a Thr Hγ1 proton of the α_0-helix (Thr 338 in Gbp2p and Thr 365 in Hrb1p) and a residue in the β_2-strand (Glu 379 in Gbp2p and Glu 406 in Hrb1p) (Figure 3.26.B). Altogether the N-terminal part of the protein blocks the access to the putative RNA-binding interface and its orientation is clearly influenced by the conformation of the linker residues between the α_0-helix and the β_1-strand. Despite the intricate network of interactions between the N-terminal region and the rest of the RRM domain, the stability of these new elements seems to be lower than the rest of the domain as judged by the rapid disappearance of its backbone amide groups in D_2O (^{1}H-^{15}N HSQCs).

3.2.4.3. Inter domain contact analysis

Independent domains can be involved in intramolecular interactions in the context of a larger construct (*i.e.* full-length). To evaluate this possibility in the case of Gbp2p/Hrb1p, two different approaches were performed. Previous data suggest that interactions between RRM1 and RRM2 might explain the stabilization observed in the tandem RRM12 (chemical-physical characterization). X-ray crystallography was attempted to obtain the structure of RRM12 tandem. However, after several crystallization trials, screening a great variety of conditions (Appendix 6), no crystal formation was detected in any case. As an alternative approach, the ^{1}H-^{15}N HSQC spectra of all constructs were compared and the chemical shift differences mapped onto the structures of RRM1 and RRM2, providing information about which amino acids were perturbed and so, putatively involved in inter domain interactions.

Chemical shift perturbations were obtained for the ^{1}H-^{15}N HSQC spectra pairs RRM123/RRM12 and RRM123/RRM3 and for the RRM12/RRM1 and RRM12/RRM2, for both proteins (Figure 3.27). Chemical shift averaging was calculated as described in 2.12.4 section.

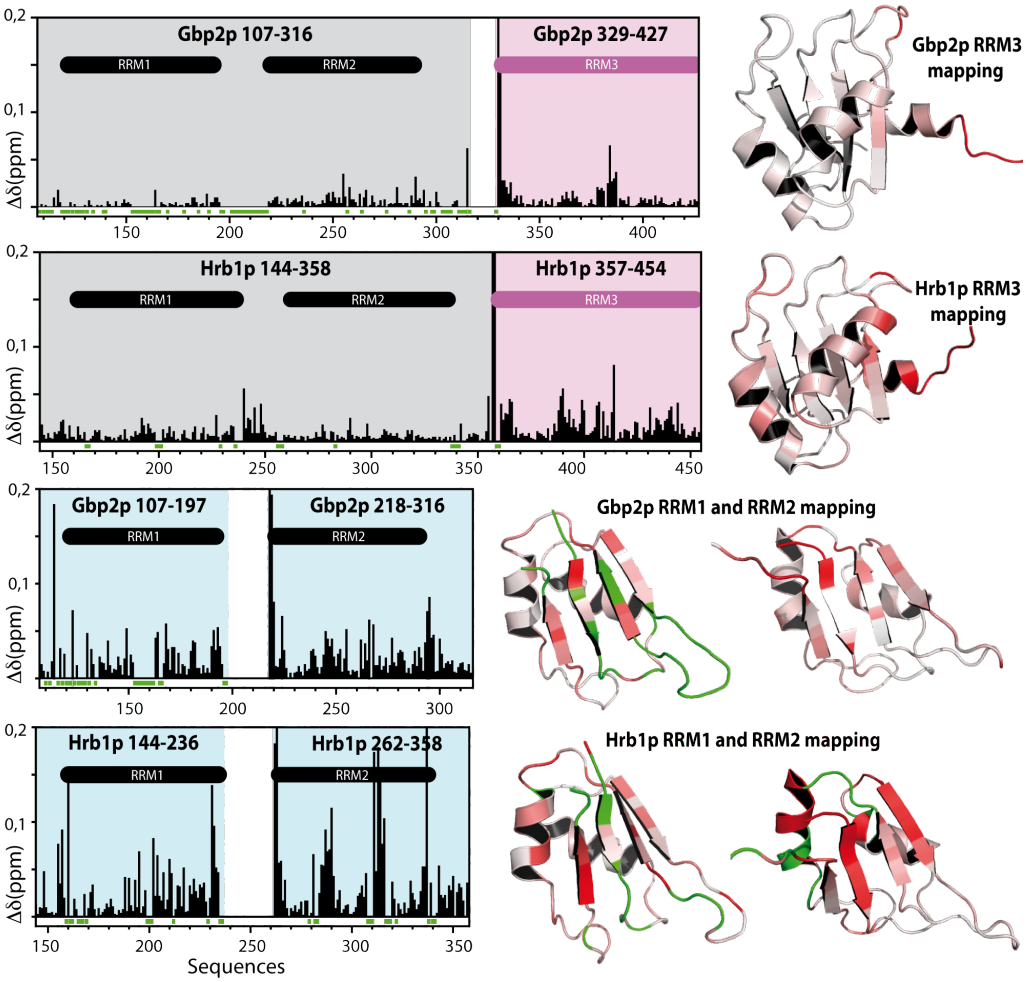


Figure 3.27. – Histograms displaying the averaged chemical shift changes of Gbp2p and Hrb1p residues when RRM123 is divided into RRM12 tandem and RRM3 (up) and in this case of RRM12 tandem split in RRM1 and RRM2 constructs (down).

Within the RRM123 context, the greatest chemical shift differences correspond to the RRM12 C-terminal and RRM3 N-terminal boundaries, attributable to the different chemical environment of this region once RRM123 is split; but additionally, some subtle differences could be detected. Mapping the perturbations on RRM3 domains two different regions were identified: a common zone surrounding the N-terminal boundary for both Gbp2p and Hrb1p

RRM3 domains (probably affected by the splitting) and the α-helixes region of Hrb1p domain but not for Gbp2p one. Curiously, some of the most perturbed amino acids in the α-helixes of Hrb1p are not conserved in Gbp2p: Glu 397 and Thr 398 in Hrb1p in similar positions of Gly 370 and Pro 371 of Gbp2p (α_1-helix) or Val 326, Asp 327 and Val 331 in Hrb1p substituted in Gbp2p by 399 Leu, 400 Val and Phe 404. In addition to this perturbation in Hrb1p RRM3, there is another region with higher chemical shift variations in the linker between RRM1 and RRM2. This region could be transiently interacting with the RRM3 domain by the α-helixes interface producing the observed perturbations. In Gbp2p this linker could not be assigned thus there is no information about its contacts, but comparing Hrb1p and Gbp2p linker sequences there are great differences, supporting the idea that the interdomain contacts observed in Hrb1p are not conserved in Gbp2p.

In the case of RRM12 tandems, their splitting entails bigger chemical shift changes than in the RRM123 dissection. The larger chemical shift deviations, excluding the RRM1 C-terminal and RRM2 N-terminal regions, are clustered along the exposed face of the β-sheet and the α_2-helix; although the unassigned regions prevent its total confirmation. The magnitude of these chemical shifts perturbations is more drastic in Hrb1p tandem.

A final analysis of the global arrangement was made in the case of Hrb1p protein: an exhaustive study of the proton/deuterium exchange was performed for the RRM123 constructs in order to determine the stability of the three RRMs in the large construct compared with those experiments carried out in the previous structural study (Figure 3.22). A liophylized sample of Hrb1p RRM123 protein was resuspended in D_2O and a set of ^{1}H-^{15}N HSQC experiments were acquired at different times to follow H/D exchange in amide signals. The results (Figure 3.28) show that at the beginning of the experiment, signals of all RRMs corresponding, by spectra comparison, to secondary structure elements were protected. While time was passing, many peaks disappeared and after 12 hours only peaks corresponding to RRM3 domain were observed (Figure 3.28).

This result confirms that the third domain has an additional stability (probably due to its compact structure) and also suggests that RRM2 domain increases its stability being within the large construct (as previous results with single domain construct showed that all signals were lost at the beginning of the experiment).

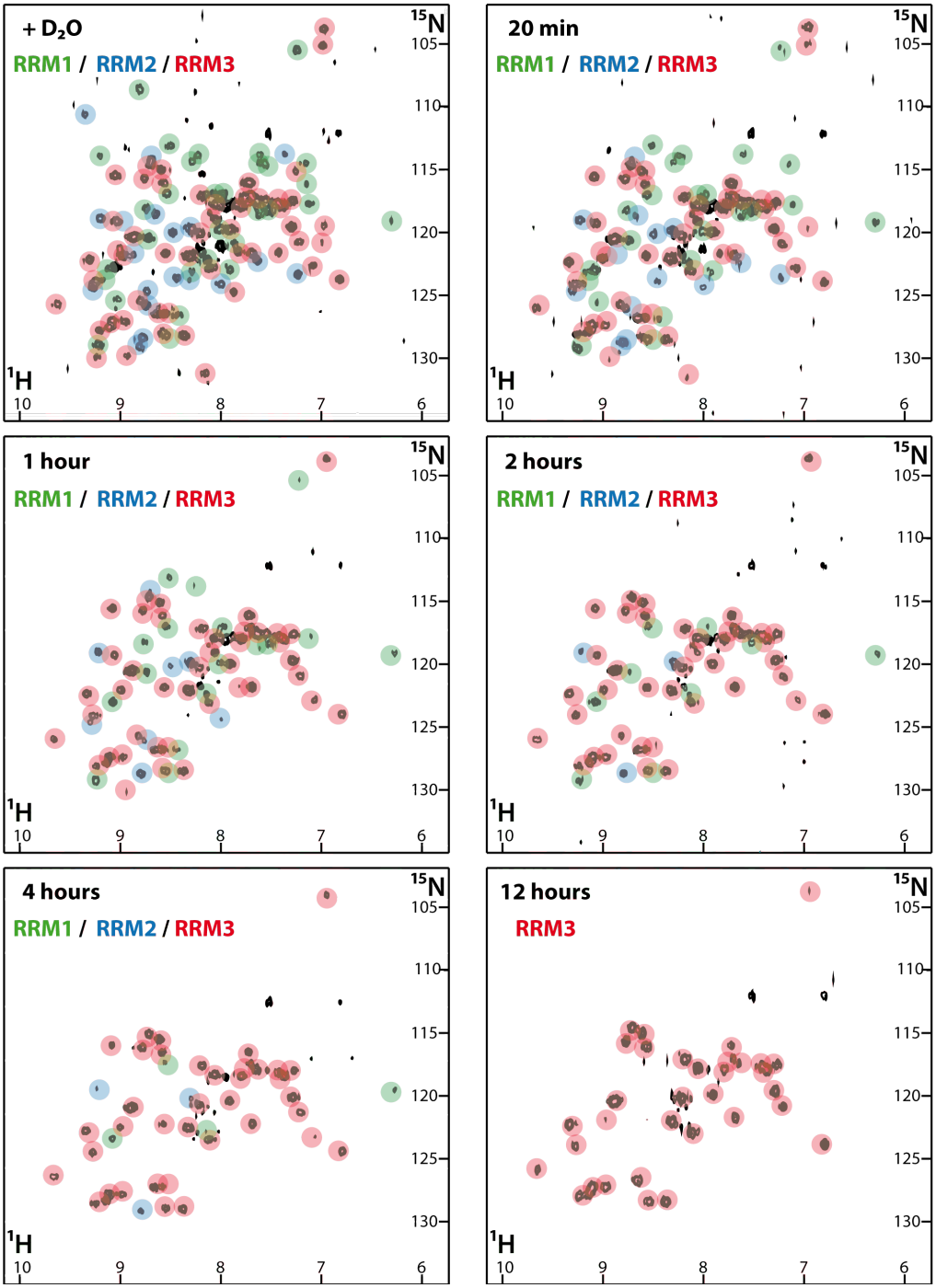


Figure 3.28. – ^{1}H-^{15}N HSQC spectra of Hrb1p RRM123 protein in D_2O acquired at different times. Peaks are highlighted in different colours depending on the domain they belong: RRM1-green, RRM2-blue and RRM3-red.

3.2.4.4. Conclusions of the structure analysis

In summary, the structural analysis reveals some essential clues, which help to guide the next binding and functional experiments.

The most important open question is about the possible role of the novel RRM3 domains, which presents a putative nucleic acid binding interface blocked by the N-terminal α_0-helix. It is important to determine if this domain is capable to bind nucleic acids and, if so, which interface it would use for it.

On the other hand, RRM12 tandem structural study entailed many difficulties, but nevertheless several data suggested that either there is an interaction between RRM1 and RRM2 that stabilizes the folding or there is a truncation effect that compromises their stabilities, thus the nucleic acid binding studies of these tandems should be done considering both RRMs in a same unit to avoid these effects.

All these aspects and others are covered in next two sections, one about nucleic acid interaction and the last one centred in *in vivo* functional studies and protein-protein interactions.

3.2.5. Nucleic acid recognition

Gbp2p and Hrb1p are described as poly (A) + RNA binding proteins (Windgassen and Krebber, 2003) that participate in mRNA transcription and export, therefore their interaction with nucleic acids is supposed to play a central role on their function. Up to now, there are two main published experimental works studying different aspects of nucleic acid interactions with these proteins: Gbp2p was described as a ssDNA binding protein involved in telomere location maintenance (Hiraga et al., 2008) and high affinity RNA sequences have been recently identify by SELEX for Hrb1p and Gbp2p (Riordan et al., 2011). This thesis progresses on such knowledge by performing several experiments in order to get more information about the affinity and selectivity of both proteins for different RNA/DNA probes. In contrast to previous data, this study dissects the contribution of different domains, regions and amino acids that participate in RNA/DNA recognition in both proteins.

Based on those previous works (Lin and Zakian, 1994; Riordan et al., 2011), the following nucleic acid probes were selected for this study (Table 3.7): some yeast telomeric DNA and RNA probes were utilized (TG-43 and derivatives) and also, a consensus RNA selected by combination of the Hrb1p and Gbp2p SELEX proposed sequences.

Nucleic acids		Sequence	Based on
Telomeric DNA sequences	TG-43	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG	(Lin and Zakian 1994)
	YG3	TGTGTGGGTGTGTGGGTGTGTGGGTGTG	(Lin and Zakian 1994)
	TG-36	GTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG	Fragment of TG-43
	TG-30	GTGTGTGGGTGTGGTGGGTGTGTGGGTGTG	Fragment of TG-43
	TG-24	GGGTGTGGTGGGTGTGTGGGTGTG	Fragment of TG-43
	TG-16	TGGGTGTGTGGGTGTG	Fragment of TG-43
	TG-12	GGGTGTGGTGGG	Fragment of TG-43
	TG-43m1	GTGGTGTGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG	Mutant of TG-43
	TG-43m2	GTGGTGGGTGTGTGTGTGTGGGTGTGGTGGGTGTGTGGGTGTG	Mutant of TG-43
	TG-43m3	GTGGTGGGTGGGTGTGTGTGTGTGTGTGGTGGGTGTGTGGGTGTG	Mutant of TG-43
	TG-43m4	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGTGTGTGTGGGTGTG	Mutant of TG-43
	TG-43m5	GTGGTGGGTGGGTGTGTGTGGGTGTGGTGGGTGTGTGTGTGTG	Mutant of TG-43
RNA probes	UG-36	GUGGGUGUGUGUGGGUGUGGUGGGUGUGUGGGUGUG	Similar to TG-36
	SELEX	UUGGUGUU	(Riordan, Herschlag et al. 2011)

Table 3.7. – RNA and DNA probes used in this work derived from previous published data.

RRM domains are versatile platforms for nucleic acids recognition (Daubner et al., 2013; Maris et al., 2005), thus, it is convenient to made a preliminary analysis of the canonical RNA binding faces present in each RRM using the structural and sequence-based data. The alignment of the RNP1 and RNP2 sequence motifs of the six RRMs of Gbp2p and Hrb1p (Figure 3.29) reveals that neither of them have all the important residues for RNA binding conserved. RRM1 domains in both proteins, which have quite similar RNP1 and RNP2 sequences, only conserve one of the three aromatic residues involved in RNA bases recognition (locus 13 on Figure 3.23). Gbp2p and Hrb1p RRM3 domains have identical RNP1 and RNP2 and, as the previous case, they also lack two of the three important aromatic residues; in addition the positive charged residue is missing. Gbp2p RRM2 has RNP1/RNP2 motifs more similar to canonical ones (it only lacks one aromatic residue in the RNP1 sequence). In contrast, its *alter ego* in Hrb1p is the most different RRM across the whole set, lacking all the important aromatic residues and the positive charged one.

From the structural point of view it is convenient to mention that RRM1/RRM2 RNA binding interfaces are exposed and in theory capable to interact with RNA, whereas in RRM3 domains this canonical binding interface is blocked by the N-terminal extension. Nevertheless, other nucleic acids binding interfaces, different from the β-sheet, have been described for RRMs (Clery et al., 2008; Daubner et al., 2013), thus the RNA binding capability of these domains can not be exclusively inferred from this sequence analysis.

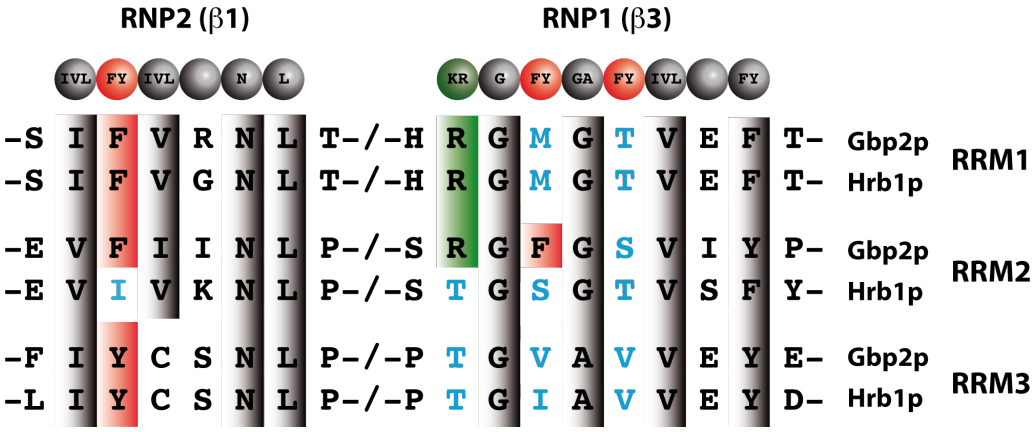


Figure 3.29. – RNP1 and RNP2 sequences of the RRMs of Gbp2p and Hrb1p, showing in blue those positions important for RNA binding and not conserved along different RRMs.

3.2.5.1. Recognition of telomere-derived DNA (RNA)

3.2.5.1.1. f-EMSA studies

Electro Mobility Shift Assay (EMSA) is a simple method that gives information about the nucleic acid binding of a protein using polyacrylamide gel electrophoresis in native conditions (see 2.7.4 section). However the sensitivity of this technique depends on the resolution of the gel (ability to separate bound and free species) and the performance of the detection method. Radiolabelling is the most sensitive detection approach, but it is dangerous and requires special licences for manipulation, thus fluorescent DNA/RNA probes are used as alternatives (f-EMSA). In this case, different fluorescent labelled DNA probes (attaching a fluorescein molecule in the 5' end of the DNA) were used to explore the binding capabilities of different Hrb1p/Gbp2p constructs. The DNAs were mixed at a constant concentration of 5 μM for each lane and protein concentrations varied from 0 to 20 or 100 μM (higher concentrations were used for single RRM constructs) (Figure 3.30).

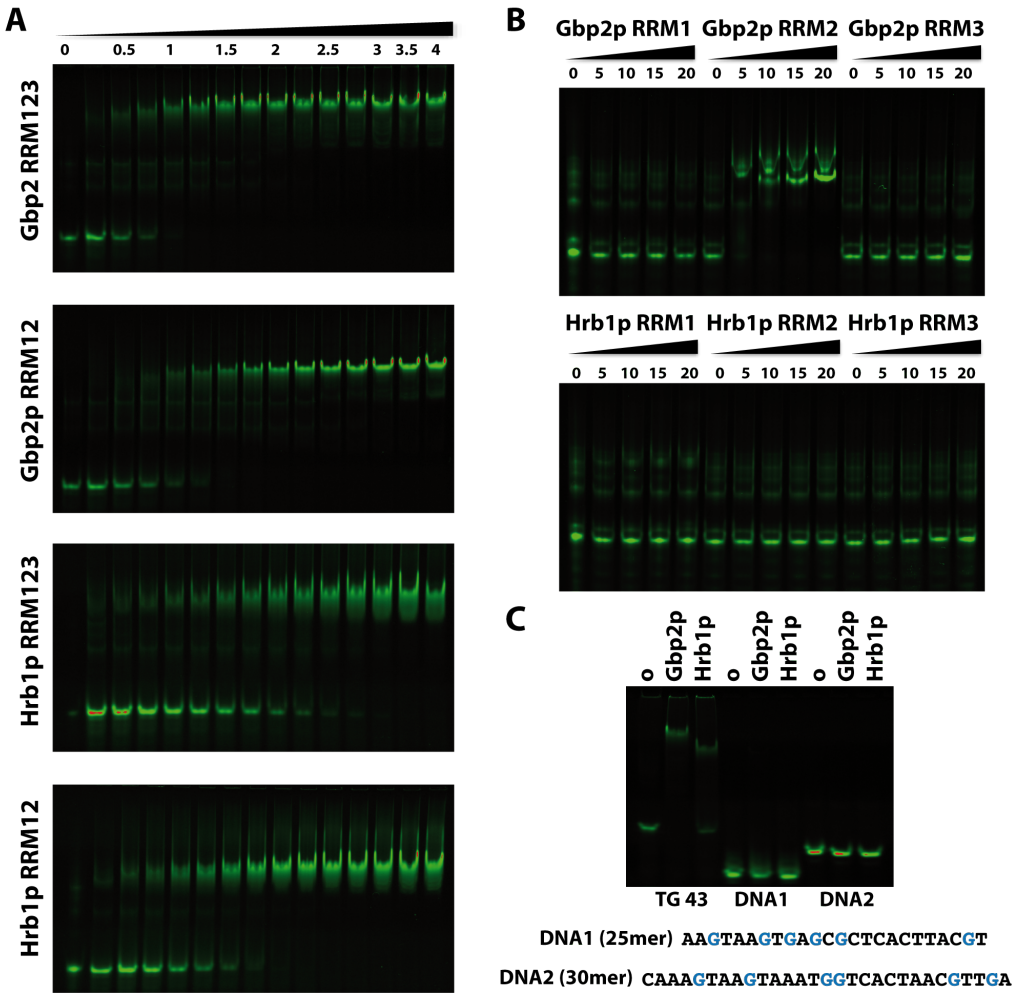


Figure 3.30. – Protein:DNA titrations monitored by f-EMSA. **A**) Fluorescent-labelled TG-43 (50 pmol) DNA titrated with increasing amounts of Gbp2p (and Hrb1p) RRM123, RRM12 constructs. Values on the top of the gel indicate DNA:protein ratios. **B**) Equivalent experiment carried out with Gbp2p/Hrb1p individual RRM constructs. TG-43 levels were identical as in the previous assay. **C**) DNA specificity assay performed with Gbp2p and Hrb1p RRM123 construct and different fluorescent DNAs. The amount of DNA and proteins was maintained fixed (5 μM and 20 μM respectively in 10 μl loaded).

The f-EMSA results show that RRM123 and RRM12 constructs of both proteins induced band shift of the TG-43 DNA probe. Despite having identical amounts of labelled DNA, the band corresponding to the free probe (down) persists to higher protein:DNA ratios in Hrb1p constructs than in Gbp2p ones, suggesting that Gbp2p binds with higher affinity (Figure 3.30.A). The f-EMSA shifts are similar when comparing two constructs of the same protein, suggesting that RRM3 domains are not necessary for DNA binding; perhaps Hrb1p RRM12

seems to bind a bit better than the RRM123 construct. Further dissection (Figure 3.30.B) reveals that, in general, RRM single constructs are not capable to induce nucleic acid shifts under these conditions. The exceptions to this rule are: Gbp2p RRM2, capable to induce band shifts by itself, and Hrb1p RRM1 that produces a slight shift. It should be noted that the protein concentrations used for single RRM assays are much higher than in larger constructs, evidencing that the affinity must be lower.

To test the selectivity of these proteins for TG rich sequences, their binding to TG-43 probe was compared with two unrelated sequences (DNA1 and DNA2) by an additional f-EMSA experiment. It was found (Figure 3.30.C) that Gbp2p and Hrb1p RRM123 constructs bind TG-43 DNA, but not the other two probes (at the same conditions). This shows that the interaction has some specificity.

3.2.5.1.2. Fluorescence anisotropy studies

Fluorescence anisotropy experiments were performed in an attempt to quantify the affinity of different constructs of Gbp2p and Hrb1p. TG-43 (and TG-36) fluorescent proves were used to get binding titrations curves for Gbp2p RRM123, RRM12, Hrb1p RRM123 and Hrb1p RRM12 constructs. The binding affinities of shorter DNA probes and RNA were estimated using competition assays. Experimental conditions were equivalent for all proteins: 20 nM DNA and protein concentration from 1 nM to 20 μM (further details in 2.10.2 section). The obtained curves (Figure 3.31) show typical titration isotherms, which are well fitted to the simplest 1:1 model (Table 3.8). At higher protein concentrations (> 5 μM), the curves exhibit a second process (linear-like), probably caused by protein aggregation.

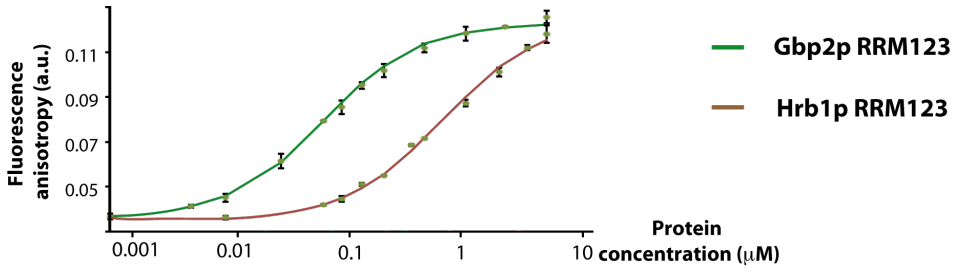


Figure 3.31. – Anisotropy titration data of RRM123 constructs of Gbp2p and Hrb1p and the fitting curve of each one. Protein concentration was expressed in logarithmic scale and the fluorescence signal in arbitrary units.

Fluorescence anisotropy titration curves do not provide reliable information about the stoichiometry of the process. However, for simplicity, the 1:1 complex model was selected to evaluate the apparent affinity of both proteins. This model would still be valid for higher stoichiometries, considering the oligomer (*e.g.* the protein) as the active specie in the equilibrium and correcting the concentration accordingly. In contrast, the kinetic model does not consider other processes that could occur simultaneously such as the observed protein aggregation or the formation of some DNA structures (as the telomeric DNA could stay linearized or folded in different quadruplex conformations); these lateral equilibria will further complicate the analysis of the obtained isotherms. With these cautions in mind, the K_D values obtained in the fitting (Table 3.8) are in the range of nanomolar that correspond to a strong protein-DNA interaction. In addition, Gbp2p presents clearly higher affinity for TG-43 DNA than Hrb1p and, in the case of Gbp2p, RRM123 and RRM12 constructs bind with similar affinity whereas the Hrb1p RRM12 construct seems to bind better than Hrb1p RRM123.

Protein	DNA	ΔG (kcal/mol)	K_D (nM)
Gbp2p RRM123	f-TG-43	9.7 ± 0.1	59 ± 10
Gbp2p RRM12	f-TG-43	9.8 ± 0.1	50 ± 9
Hrb1p RRM123	f-TG-43	8.2 ± 0.1	777 ± 267
Hrb1p RRM12	f-TG-43	8.8 ± 0.2	289 ± 96

Table 3.8. – Energetic parameters of the interactions between various constructs of Gbp2p and Hrb1p with TG-43 DNA obtained by fluorescence anisotropy.

In an effort to define the binding sequence more precisely, competition experiments were performed adding shorter DNA probes derived from TG-43 (Table 3.7) at a constant concentration of 20 mM to the original experimental setup. According to these (Figure 3.32), the competitor efficiency is proportional to its length; the longer the DNA competitor the higher the effect. This behaviour suggests that rather than binding to a high affinity region in the probe, the proteins can bind to multiple sites with similar affinity along the TG sequence. In addition, these results indicate that the complex could not have 1:1 stoichiometry and an undetermined number of protein molecules are simultaneously bound to the same DNA probe.

All these results evidence that the recognition of telomeric DNA sequences by Gbp2p is a complex process that cannot be studied using a single technique (*i.e.* fluorescence anisotropy) but rather require a multiple approach. Primary on this is to establish the conformational properties of TG-43 and to which species (single strand or quadruplex) do the proteins bind.

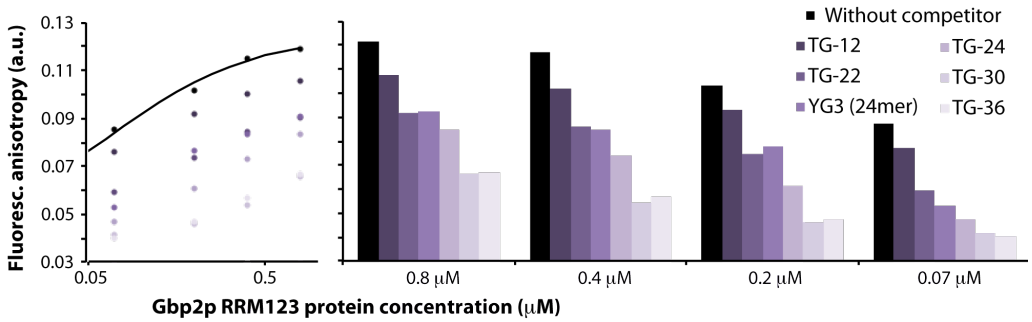


Figure 3.32. – A) Fluorescence anisotropy titration curves of TG-43 DNA (20 nM) with increasing amounts of concentrations of Gbp2p RRM123 and in the presence of various DNA competitors (at 20 mM). B) The same data is represented as a bar chart for each protein concentration tested.

3.2.5.1.3. Circular dichroism

The first evidence of Gbp2p function suggested a possible role in telomere binding, maintenance and localization (Konkel et al., 1995; Lin and Zakian, 1994). Recently, some *in vivo* experiments localize this protein in the telomers (Hiraga et al., 2008) and somewhat reinforce that primary role described long time ago. The sequence of telomeric DNA (TG-43) suggests that it can potentially form linear or quadruplex arrangements. Thus, it is important to determine if these two conformations exist and, if so, to which of them Gbp2p does bind preferentially.

The native PAGE analysis of TG-43 and its shorter version TG-36 shows that the two molecules exist as a mixture of species (Figure 3.33.A). In the presence of K^+ ion, the TG-43 can form more than 4 different species, of which only two (the two upper bands) are present in the absence of K^+. TG-43 has 5 GGG motifs, thus the extra fast migrating bands that appear after annealing in the presence of potassium are likely different forms of quadruplexes (structurally more compact). Specific mutations on these five GGG motifs of TG-43 (Table 3.7) reduce the heterogeneity of the system and the surviving quadruplexes bands match with the different bands observed in the TG-43 sample (Figure 3.33.A-right). In the absence of potassium (Figure 3.33.A-left), the upper band is the predominant form in all the cases and likely correspond to the linear conformer.

These results are further supported by circular dichroism experiments that identify the characteristic bands of quadruplex structures in the near UV region (parallel quadruplex present a band at 260 nm and antiparallel one at 290 nm) (Kypr et al., 2009). The shorter

version (TG-36) of telomeric DNA, that contains 4 guanine triads instead of 5, was used to reduce the conformational heterogeneity of the original DNA probe.

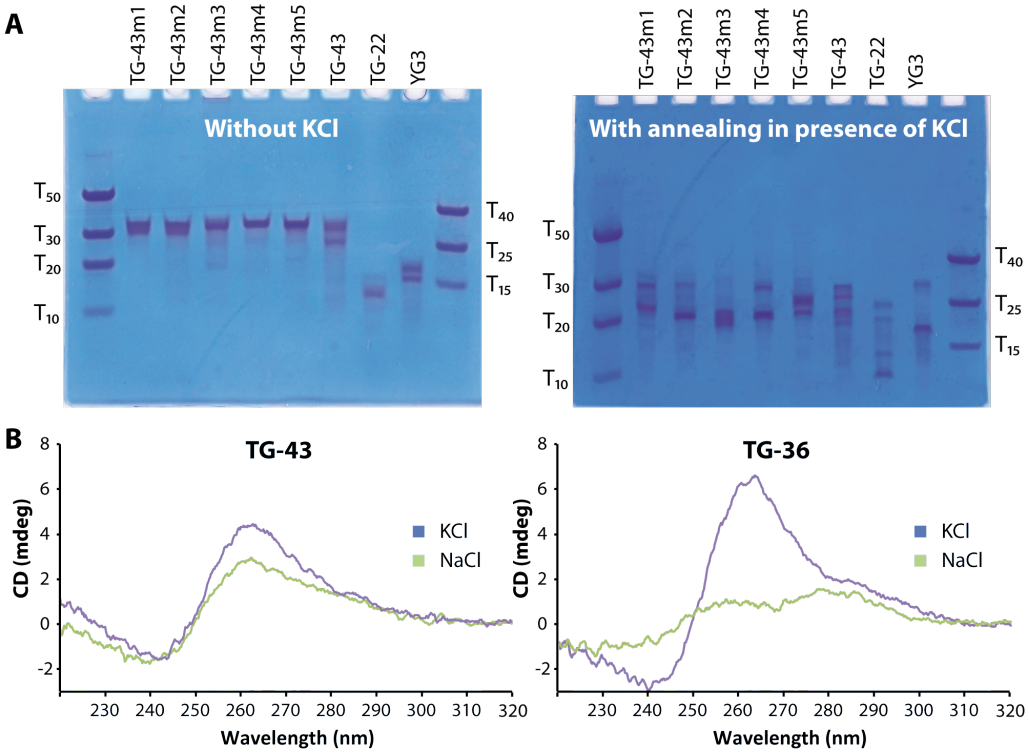


Figure 3.33. – Telomeric G-strand TG-43 (and its derivate TG-36) can form quadruplex-like structures. **A**) Native PAGE of TG-43, shorter DNA versions TG-22 and YG3 and some TG-43 GGG triad mutants, without KCl (left) and annealed in the presence of potassium (right). **B**) Near UV circular dichroism spectra of both DNA probes at 7.5 µM in 150 mM KCl (violet) and in 150 mM NaCl (green) buffers (in 20 mM potassium or sodium phosphate at pH 6.5 respectively).

As expected, these DNA structures are stabilized by coordination to different cations, typically K^+ (Bochman et al., 2012). In Figure 3.33.B, the appearance of a band at around 265 nm in de presence of 150 mM KCl evidence the existence of these DNA structures in both probes. In addition, a spectra comparison between KCl and NaCl buffers shows that K^+ ions stabilize better the quadruplex structure; although using Na^+ in TG-43, and in lesser extent in TG-36, some amount of quadruplex is present. This residual presence of quadruplex might be explained by the experimental difficulty to deplete K^+ concentration to near-to-zero values in the DNA samples (it is a common counter ion introduced during the synthesis).

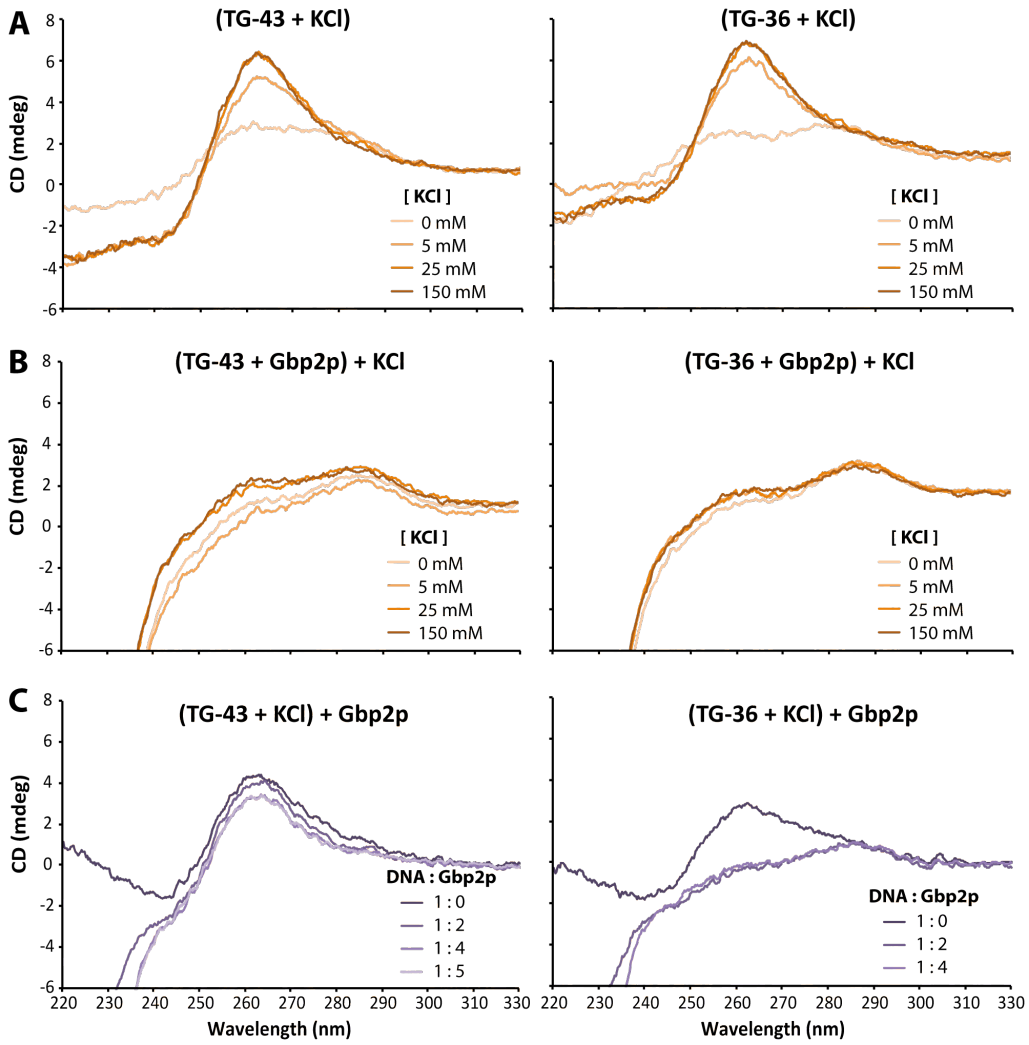


Figure 3.34. – Different near UV CD spectra showing: A) A titration of TG-43 (left) and TG-36 (right) DNAs (at 7.5 μM) with KCl. B) The same titration with KCl but to a previously Gbp2p RRM12 – DNA preformed complex. C) A titration of Gbp2p RRM12 protein (at different concentrations) to different preformed quadruplex of DNA (7.5 μM) in KCl buffer.

To investigate the effect of protein binding in the CD spectra, different binding assays using Gbp2p RRM12 construct were performed. As the conformational state of the DNA can be modulated by the presence of different cations, the mixtures of the three components (DNA, protein, cation) were prepared, finding that the results depended on the order followed to mix the components. When the quadruplex is pre-formed (DNA in presence of potassium) very few amount of TG-43 quadruplex is denatured when the protein is added (Figure 3.34.A and

C). In contrast, TG-36 quadruplex with K^+ and TG-43 quadruplexes formed in the presence of Na^+, are easily unfolded by the addition of Gbp2p RRM12 (probably due to their lower stability than TG-43 structures formed with KCl). It is worth mentioning that, some antiparallel conformation seems to remain as the band at 290 nm is not disturbed. By contrast, if the DNA – protein complex is pre-formed, when the KCl is added there is no appearance of quadruplex structures (Figure 3.34.B), evidencing that Gbp2p protein binds DNA in a way that inhibits quadruplex formation. An intermediate situation happens when DNA is added to a solution of Gbp2p protein in KCl 150 mM buffer, some DNA forms quadruplex and the rest, presumably protected by Gbp2p, does not.

Taken together, these results suggest that Gbp2p binding to TG-43 (TG-36) linear DNA (probably the major form in the absence of potassium) might form kinetically trapped complexes with slow dissociation kinetics (high values of activation energy, Figure 3.35), thus the concentration of free DNA able to form quadruplexes is substantially depleted. The quadruplex structure does not seem to be required for Gbp2p binding and DNA-protein interaction probably involves the recognition of nucleotide bases in a manner that competes with guanine-guanine contacts in the tetrads.

The fact that Gbp2p is still able to bind to TG-43 in the presence of KCl, without disrupting the non-binding conformation (the quadruplex) might be rationalised by the existence of species having long ssDNA regions to which Gbp2p would still be able to bind. The slow dissociation kinetics of the DNA quadruplexes would explain why Gbp2p is not capable to melt them (Figure 3.35).

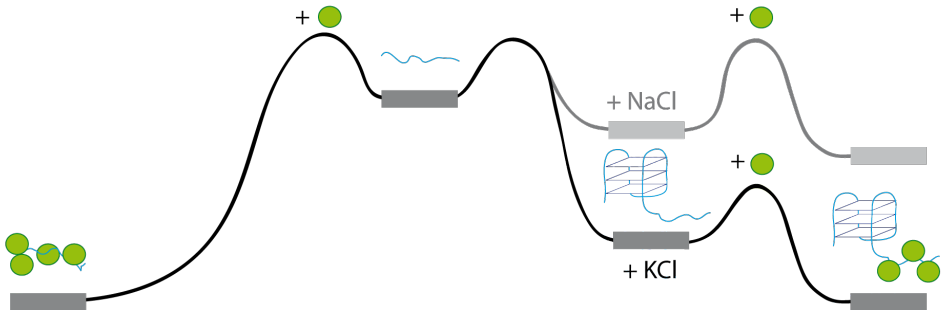


Figure 3.35. – Schematic representation of energetic values of TG-43 DNA in binding and quadruplex formation.

3.2.5.1.4. NMR titrations

NMR was used as the last experimental procedure in the analysis of the telomeric nucleic acid binding of Gbp2p/Hrb1p proteins. The method consisted on NMR titration assays of several Gbp2p and Hrb1p constructs with some DNA/RNA probes. Titrations using TG-43 DNA were monitored in the ^{1}H-^{15}N HSQC spectra of Gbp2p/Hrb1p RRM123 constructs (Figure 3.36.A). The results showed the disappearance of all signals corresponding to RRM1 and RRM2 domains for both proteins while the RRM3 signals were hardly perturbed (see 3.2.5.3 section).

In the following experiment, TG-43 DNA binding was analysed with a similar approach for single RRM constructs of both proteins. The ^{1}H-^{15}N HSQC spectra of Gbp2p RRM1, RRM2 and Hrb1p RRM1 experience the same global signal disappearance effect (Figure 3.36.B). Only the signals from the unfolded N-terminal could be observed. In the case of Hrb1p RRM2, the titration produces some slight peak shifts on the spectra but not their disappearance (Figure 3.36.C), suggesting a weaker interaction than the other RRMs.

Experiments utilizing RNA were also performed. A similar sequence to TG-36 was utilized (UG-36), this RNA was obtained in the lab by *in vitro* transcription procedures (see 2.5.3 section), and the results of the titration with RRM123 constructs showed the same pattern as DNA: all signals corresponding to RRM12 tandem vanished. Moreover the global disappearance of RRM12 resonances is still observed with DNA probes of progressively decreased size (up to TG-15 see Table 3.8).

There are several possibilities to explain the disappearance of the signals in the ^{1}H-^{15}N HSQC experiments upon titrations with TG-43. The most plausible scenario is that the RRM12 region participates in the formation of high molecular weight oligomers upon DNA binding. This would justify why the effect occurs for all the signals as the increase in the line broadening would be due to an increase in the correlation time. The alternative chemical exchange scenario cannot cause this global change, as unperturbed signals ($\Delta\delta=0$) would not experience line-broadening effects.

The formation of these high molecular weight structures occurs for DNA and RNA probes, hence it is likely that such behaviour depends on chemical moieties (guanoside bases of phosphate backbone) that are common to the two types of molecules. Nevertheless the severe broadening effect precludes the identification of the DNA binding interfaces by the chemical shift mapping method.

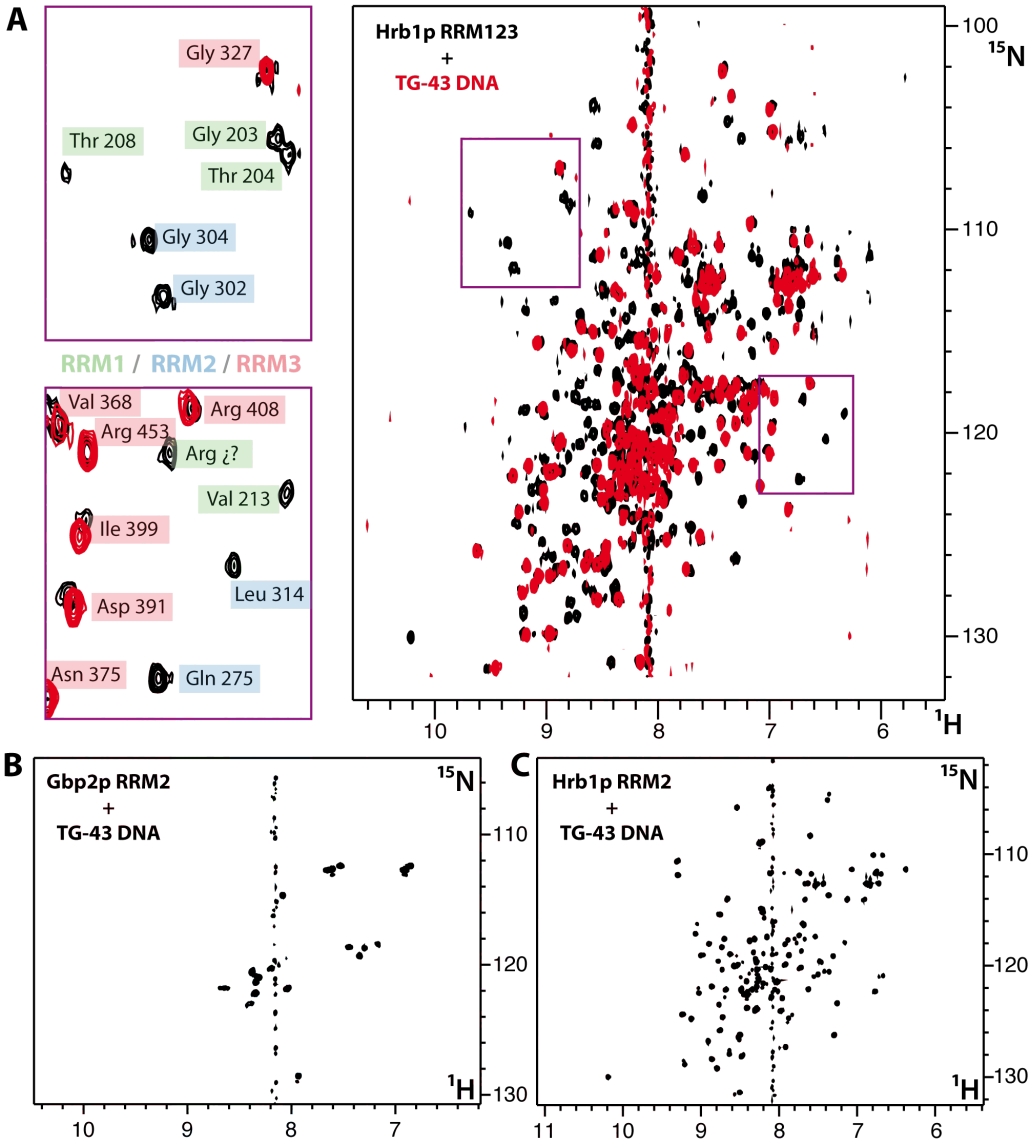


Figure 3.36. – A) ^{1}H-^{15}N HSQC spectra superposition showing the titration experiments with TG-43 DNA in Hrb1p RRM123 construct; black spectrum corresponds to free form and red one to DNA titration; two regions were amplified (left) and assigned; the labels were highlighted depending on the domain where the residues are located: RRM1-green, RRM2-blue and RRM3-red. ^{1}H-^{15}N HSQC spectra of Gbp2p RRM2 domain (B) and Hrb1p RRM2 one (C) titrated with TG-43 DNA. All experiments were made at equimolecular concentrations of protein and nucleic acid between 75 and 100 μM.

3.2.5.2 Recognition of SELEX-derived RNA

A shorter RNA probe (5'-UUGGUGUU-3') was used to obtain further insights about the nucleic acid recognition. The probe is derived from the consensus sequence identified in previously reported SELEX experiments of Gbp2p and Hrb1p proteins (Riordan et al., 2011). Just as in the case of DNA, a set of titrations was performed using this 8-mer RNA with RRM123 and single RRM constructs. In this case, signal shifts (rather than disappearance) were detected making possible to obtain an averaged value of the chemical shift change for each protein residue and hence, the different binding regions could be mapped (Figure 3.37).

The largest changes in chemical shift values were observed for Gbp2p RRM2 domain residues, mapping in the canonical β-sheet binding face and in the α_1-helix region (Figure 3.37). By contrast, Hrb1p RRM2 domain does not present big chemical shift changes except in a small group of residues around α_1-helix and β_2-β_3 strands (Figure 3.37). RRM1 domains spectra also presented chemical shift deviations during titrations and the residues perturbed by the interaction seem to belong to the canonical RNA binding face of RRMs (Figure 3.37), although the poor spectra quality in Gbp2p RRM1 avoided the complete assignment of these chemical shift deviations. The linkers between RRMs do not suffer significant perturbations, but the N-terminal part of Hrb1p RRM1 (prior to Asn 161) shows some large changes (not readily detected for Gbp2p). Finally, RRM3 domains did not present any signal perturbation, although Hrb1p RRM3 in RRM123 construct showed slight changes (see 3.2.5.3 section).

Additionally, titration experiments using other RNA probes (A_{12} and U_{12}) performed for Hrb1p RRM123 construct led to smaller chemical shift perturbations than in the UUGGUGUU RNA case, but involving the same regions described above.

3.2.5.2.1. Insights into RNA (DNA) recognition mechanism

The compendium of the experimental data obtained in this work shows that nucleic acid binding abilities of Gbp2p/Hrb1p reside on the region RRM12 and the binding mode of Gbp2 RRM2 domain presents new elements that have to be deeper analysed. The chemical shift deviation data reveals that the domain presents a secondary binding interface constituted by α_1-helix and β_2-strand; the solvent-exposed Trp 232 (on α_1-helix) is clearly perturbed (Figure 3.38.B) and it can be postulated as a central element on RNA binding. In Hrb1p RRM2 domain, this novel RNA-binding interface seems to be used as well as the conserved tryptophan residue is clearly perturbed upon titration.

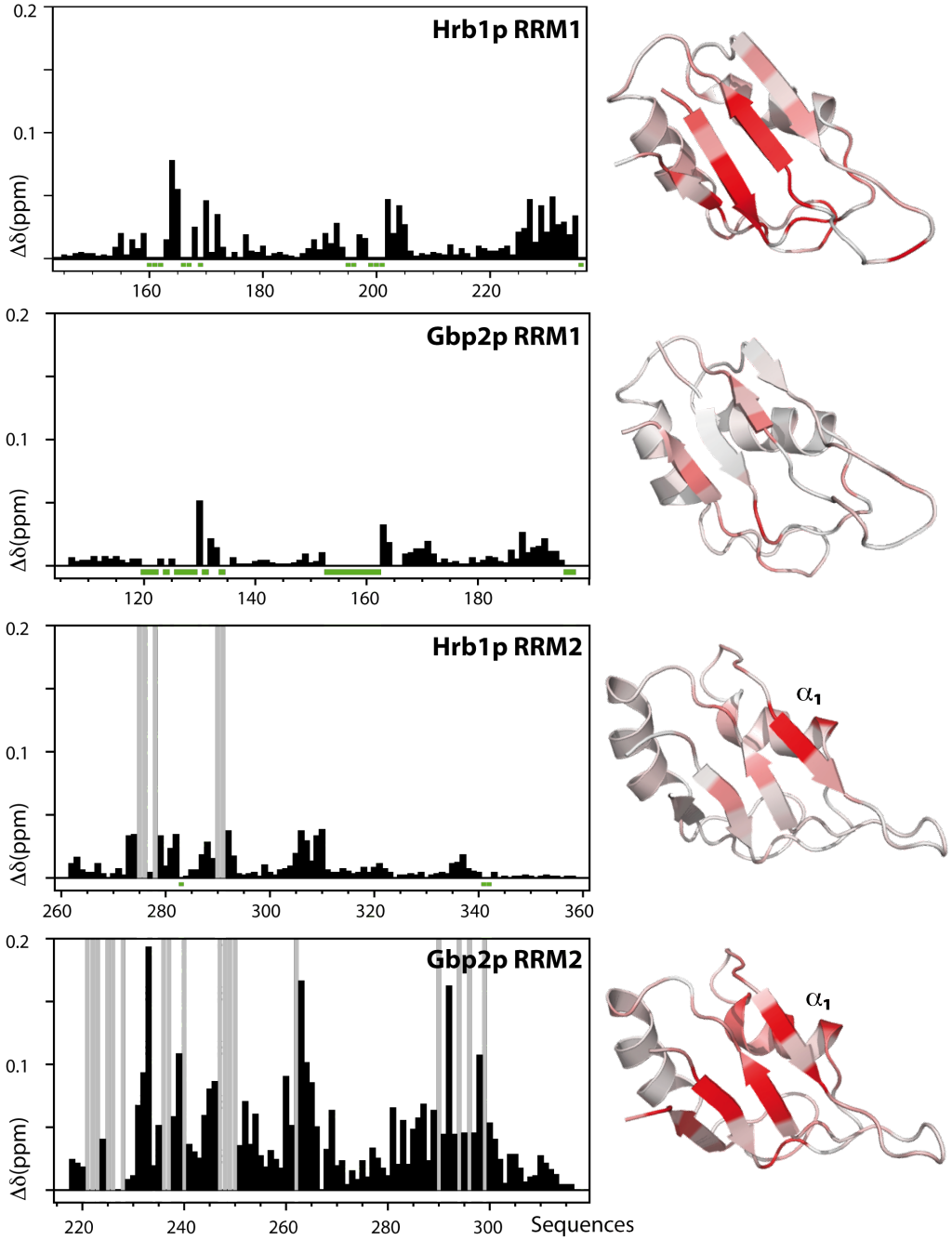


Figure 3.37. – (Left) Histograms displaying the averaged chemical shift changes of ^{1}H-^{15}N HSQC spectra during titrations of RRM1 and RRM2 domains with UUGGUGUU RNA (equimolecular mixes at 80 – 150 µM); signal loss due to binding was highlighted with grey bars and unassigned regions with green marks. (Right) Mapping of the averaged chemical shift perturbations onto the RRM structures; Gbp2p mapping is represented on Hrb1p structures.

During the writing of this thesis, it was described a new class of RRM (so called pseudo-RRM), that uses a non-canonical binding interface to recognise 5'-GGA-3' RNA specifically (Clery et al., 2013). The oncoprotein SRSF1 is a serine/arginine protein, similar to Gbp2p/Hrb1p, that participates in alternative splicing in mammals. Its second RRM is classified as a pseudo-RRM and characterized by a consensus sequence (SWQDLKD) in the α_1-helix that is conserved across species including the yeast Npl3 RRM2 domain. The second RRM of Gbp2p and Hrb1p have a nearly identical sequence (Figure 3.38.A). In the structure of SRSF1/RNA complex, Trp 134 plays a central role in the recognition of the G5, Lys 138 and Asp 139 have base specific contacts with the second G6 and Gln 135 is stacked with the aromatic ring of A7. All these elements are present in the structure of Hrb1p RRM2 (and presumably in the Gbp2p one) (Figure 3.38.C).

This particular binding mode produces distinct spectroscopic changes in the NMR spectra that have been consistently reproduced in similar proteins that contain pseudo-RRMs (Clery et al., 2013). Equivalent changes can be found when analysing the titration of Gbp2p RRM2 and Hrb1p RRM2 with 5'-UUGGUGUU-3' RNA (Figure 3.38.C). The most relevant features are the shielding of the tryptophan indole of Trp 232 in Gbp2p (and in less extension Hrb1p Trp 274) (Figure 3.38.B) and the perturbation of Gln 233 side chain (Gln 275 in Hrb1p), which are rersidues involved in stacking interactions with RNA bases in SRSF1. The backbone cross-peak of Ala 248 in Gbp2p (Ala 290 in Hrb1p) becomes severely broadened upon binding, as occurs in the equivalent amide in SRSF1 (Ala 150) that is involved in a base specific contact to the Watson-crick face of the G5. Finally, the residues Lys 138/Asp 139 that provide base specificity for G6 in SRSF1 (Figure 3.38.D) are conserved in Gbp2p/Hrb1p pseudo-RRMs.

All these evidences suggest that RRM2 of Gbp2p/Hrb1p have a similar RNA binding mode to SRSF1 and that recognise at least two consecutive guanines with high selectivity. However, the perturbations in Gbp2p extend to the classic RRM recognition surface (the β-sheet). Thus, it is possible that the two recognition modes coexist in this case (not clearly observed in Hrb1p). The cooperation between these two binding modes might result in a higher affinity in Gbp2p (larger perturbations of the Trp indole Figure 3.38.B). A hypothesis that nevertheless needs to be corroborated experimentally.

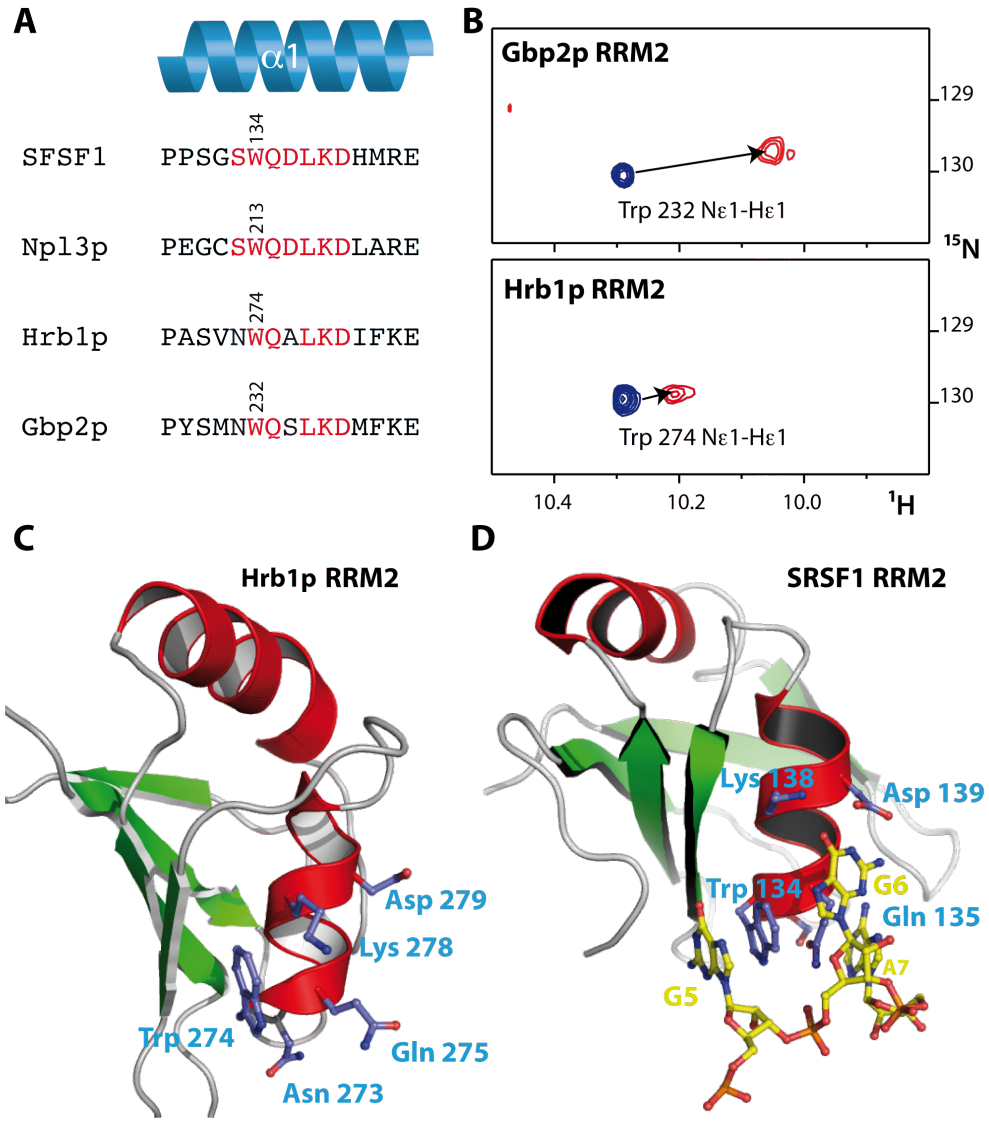


Figure 3.38. - Comparison between the RNA binding modes of pseudo-RRM domain of SRSF1 (RRM2) and Gbp2p/Hrb1p (RRM2). A) Sequence alignment of characteristic α_1-helix motifs (in red) of pseudoRRM domains of human SRSF1 and *Saccharomyces cerevisiae* hnRNP proteins. B) Characteristic chemical shift perturbation of Trp indole correlation signals in the ^{1}H-^{15}N HSQC upon binding to 5'- UUGGUGUU-3' RNA. Similar behaviour was reported for SRSF1 RNA binding (Clery et al., 2013). C) NMR structure of Hrb1p pseudo-RRM2 showing the side chains of exposed residues of the α_1 motif. D) NMR structure of the complex between SRSF1 pseudo-RRM with 5'-UGAGGAC-3' (PDB 2M8D).

In the case of SRSF1 the binding mode of the pseudo-RRM is essential for the biological function of the protein in mRNA alternative splicing (Chiodi et al., 2004; Dauksaite and Akusjarvi, 2004). In yeast, Npl3p has been shown to be involved in mRNA splicing. However, for the cases of Gbp2p and Hrb1p no such direct biological relationship has been established, but their analogy with these splicing factors deserves further investigation at the light of this new interaction data (see 3.2.6.2 section).

Also in SRSF1, the region involved in RNA binding in RRM2 is also involved in the recognition of SRPK1 (Ghosh and Adams, 2011), an SR-specific protein kinase homologous to Sky1p in yeast. Their two recognition processes are mutually excluded. The implications for the possible recognition mode of Gbp2p/Hrb1p for Sky1p and the influence on the phosphorylation mechanism are discussed later (in the 3.3. section of this chapter).

3.2.5.3 Effects of DNA/RNA binding on long-range protein-protein contacts

Previous data pointed that RRM3s are not nucleic acid binding domains. Therefore, a final set of titration experiments with the different nucleic acid probes was performed in order to confirm this lack of interaction. NMR titrations of the single RRM constructs with different probes (DNA/RNA) showed that no peak signals were perturbed, even when using other sequences in order to explore additional base selectivity (A_{12}, U_{12}, $(AU)_7$, $(CU)_5$).

Surprisingly, RRM3 peak perturbations were detected in the Hrb1p RRM123 construct titration, while the RRM3 signals were not altered in Gbp2p (Figure 3.39). Moreover, the perturbed regions coincide with those proposed in the inter domain contact analysis and as shown in Figure 3.39, the RRM123 construct titrated with RNA/DNA gives an intermediate spectrum between RRM123 construct (in RRM3 peaks) and free RRM3. All these data seem to suggest that RRM3 signal perturbation in Hrb1p is due to changes in the inter domain contacts upon nucleic acid binding rather than to direct RNA/DNA contacts.

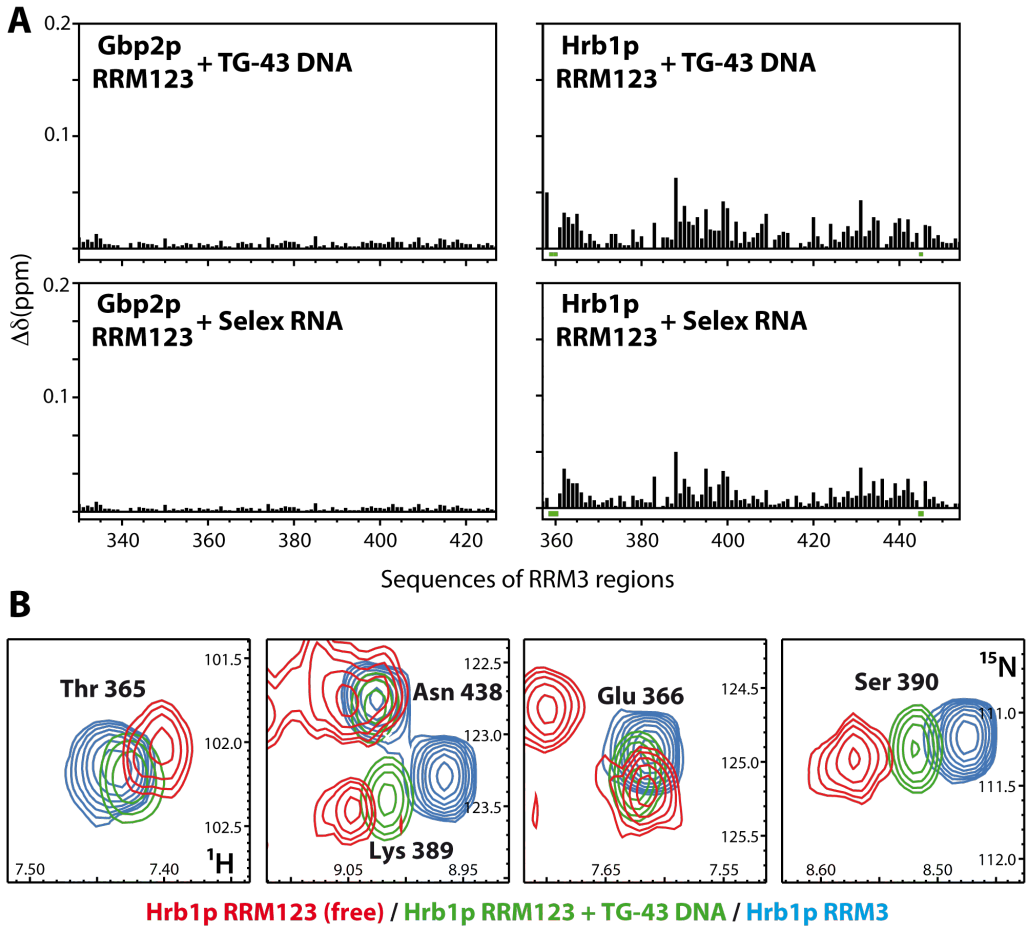


Figure 3.39. – A) Similar histogram representations to Figure 3.37, but showing the chemical shift perturbation values of RRM3 domains in RRM123 contexts for Gbp2p and Hrb1p proteins using TG-43 DNA and Selex RNA. B) Histogram extracted from Figure 3.27, showing the chemical shift perturbation of RRM3 signals between single domain construct and RRM123 one. C) Details of ^{1}H-^{15}N HSQC spectra superposition of Hrb1p RRM3 (blue), Hrb1p RRM123 (red) and Hrb1p RRM123 titrated with TG-43 (green).

3.2.6. *In vivo* functional analyses

Once the nucleic acid binding analysis was finished, a final set of experiments were performed in order to identify the role of the C-terminal RRM3 domain and also to explore the functions of Gbp2p and Hrb1p in the cell. These precise biological functions of Gbp2p/Hrb1p are not well understood and there are several evidences relating these hnRNP with cytoplasmic RNA granules, stress granules (Gbp2p) and P-bodies (Hrb1p) (Buchan et al., 2008; Mitchell et al., 2013) and mRNA translation and export (Hurt et al., 2004; Windgassen and Krebber, 2003).

These two proteins differ from Npl3p (the other SR protein in yeast) in the way they are co-transcriptional recruited to the mRNA via the THO/TREX complex (Hurt et al., 2004) and also because they are dispensable for efficient mRNA splicing (Npl3p is essential) (Kress et al., 2008; Windgassen et al., 2004). In the last part of this study, different *in vivo* functional analyses using *Saccharomyces cerevisiae* cells were carried out seeking to correlate the structural and binding data to the biological role played by these proteins in yeasts. Of particular interest is to revisit the role of these proteins in mRNA splicing (given the similarity of pseudo-RRM2 RNA binding mode with that of SRSF1) and to assign a function to the conserved RRM3 domains, which have novel structural features but surprisingly have been revealed as non nucleic acid binding domains.

3.2.6.1. Effects in cell growth

Previously published data showed that deletion of Gbp2p and Hrb1p has not a distinct phenotype upon growing the cells in rich media (Giaever et al., 2002), concluding that these proteins are not essential for yeasts. In this work an *in vivo* analysis of the global effect in cell survival was evaluated sampling several different conditions to test whether the removal of these proteins affects yeast growth under stress situations. To perform these studies, the following yeast strains were produced (see 2.6.1 section): two single deletion mutants (*gbp2Δ* and *hrb1Δ*), which were prepared by yeast homologous recombinant cloning and the double mutant strain *gbp2Δ/hrb1Δ*, which was obtained by crossing the two single mutants.

Growth rate was tested for these strains in different conditions (drop tests and liquid media) using a wild type yeast strain as control. Rich and minimal media plates supplemented with several carbon sources and also with 1M NaCl were prepared for the drop tests. These tests were performed at four temperatures (16, 25, 30 and 37 ºC) with the four strains (wild type, *gbp2Δ*, *hrb1Δ* and *gbp2Δhrb1Δ*). The results (Figure 3.40.A) showed that mutant phenotypes do not differ significantly from wild-type ones in any of the conditions assayed.

In addition to the drop test, another growth experiment was performed in liquid media. In this case only two different media were used YPD and YPD+1M NaCl. 20 ml of each media were inoculated with fresh cultures of each four strain and growth was registered by measurement of optical dispersion at 600 nm at different times. At the beginning, growth was similar for each strain, but between day 2 and 3 a subtle difference was reported: double deletion mutants reach higher OD_{600nm} than single mutants, and these strains higher values

than wild type ones (Figure 3.40.B). These data pointed that deletion of Gbp2p and Hrb1p provides some subtle benefices for cell growth in saturated cultures.

All of these results complement previously published data (Giaever et al., 2002) and confirm that Gbp2p and Hrb1p deletion does not generally affect yeast survival and cell growth in the tested conditions. Although some evidences detected in liquid media growth experiments may suggest an implication of these proteins in negative regulation, thus their deletion entails faster growth rates.

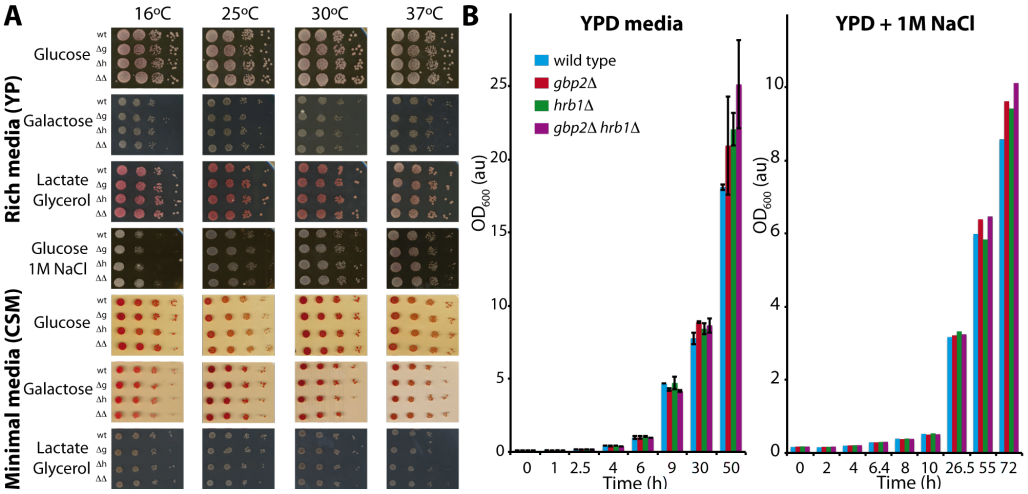


Figure 3.40. – A) Drop tests of each construct under different conditions. B) Representation of each strain growth in liquid media at different time, measuring the optical dispersion of the culture at 600 nm.

3.2.6.2. Gbp2p and Hrb1p role in mRNA splicing

The binding mode of Gbp2p/Hrp1p RRM2 described here is equivalent to that of SRSF1 pseudo-RRM (Clery et al., 2013), which in this protein is essential for its function in alternative mRNA splicing. However, there is not a solid link between Gbp2p and Hrb1p proteins with mRNA splicing process and indeed the data available rather suggest that the two proteins are not necessary for this function (Kress et al., 2008). Nevertheless, the presence of the pseudo-RRM opens the question weather it is involved in splicing regulation (as described for these type of RRM) or not. To clarify if these proteins are related to mRNA splicing, experiments comparing the activity of an intron-containing reporter in the different *gbp2Δ* and *hrb1Δ* backgrounds were carried out. To perform that, β-galactosidase reporter assays were made in order to test splicing and expression levels using different plasmids in

which β-galactosidase gene is fussed to several intron-containing sequences. The experiment consisted on a comparison of the β-galactosidase activity of these different reporters in protein extracts coming from wild-type and mutant (*gbp2Δ/hrb1Δ*) strains.

A first analysis was carried out using three different plasmids: a negative control plasmid without β-galactosidase gene, a positive control plasmid that incorporates the reporter gene without any intron and the splicing test plasmid where the β-galactosidase gene is fused to the 5-end of an intron-containing segment of the RP51 gene (YML024W in the standard yeast nomenclature) (Figure 3.41.A). The experiment procedure consisted in growing each strain transformed with the different plasmids, inducing the reporter expression by addition of galactose and obtaining each protein extract to perform a simple β-galactosidase assay using ONPG as substrate (see 2.6.2 section). The results of these first assays (Figure 3.41.B) showed that the β-galactosidase expression levels were different for each plasmid. There is a small but significant difference in activity between double deletion mutant and wild type strains within each type of plasmid (clearly seen for positive control), which is indicative of a higher β-galactosidase activity in *gbp2Δ/hrb1Δ* strain than in wild type one.

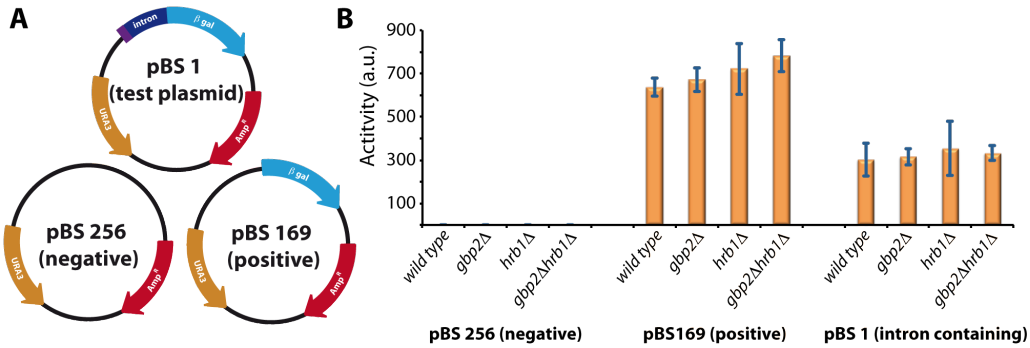


Figure 3.41. – A) Representation of plasmids used in the β-galactosidase tests. B) Calculated β-galactosidase activity for each strain (wild type, *gbp2Δ*, *hrb1Δ* and *gbp2Δ/hrb1Δ*) transformed with each plasmid (negative control, positive control (β-galactosidase) and intron-containing β-galactosidase).

The same procedure was used for several different plasmids containing mutations in the intron sequence or other different introns, but in this case only wild type and the double mutant strains were assayed. The results confirmed that in all cases the double mutant strain showed slightly more reporter activity (Figure 3.42).

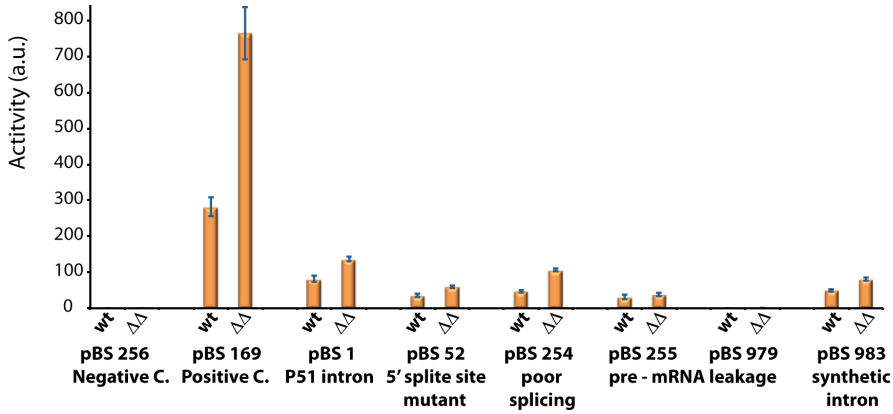


Figure 3.42. – Results of β-galactosidase assay using different intron containing plasmids. A – negative control plasmid; B – positive control plasmid incorporating native β-galactosidase gene; C – reporter plasmid containing β-galactosidase gene fused to RP51 intron; D, E, F and G – mutants of plasmid C in different regions that promote poor splicing and/or pre-mRNA leakage to the cytoplasm; H – similar plasmid to C but containing a synthetic intron.

As the positive control plasmid (not affected by splicing) experiences the same effect than the intron containing ones, it seems unlikely that changes in β-galactosidase expression levels between wild-type and mutant backgrounds are correlated to an active role of Gbp2p/Hrb1p in splicing, but rather to a more global effect, yet to be determined.

Although the global tendency found from comparison of these two experiments is conserved and the double mutant strain expresses more reporter than the wild type in all cases, the magnitude of this activity difference is not constant, suggesting that some different conditions between assays may affect the results. To test that, a set of β-galactosidase experiments was performed changing growing conditions and using the positive control plasmid and the original and double mutant strains. Different induction times and temperatures and also different growing media were tested (Figure 3.43). In all cases the expression seems to be higher in the double mutant strain than in the wild type except at temperatures lower than 30ºC, when the difference is reverted (Figure 3.43.A). The media (Figure 3.43.B) and the different induction times did not change substantially this pattern.

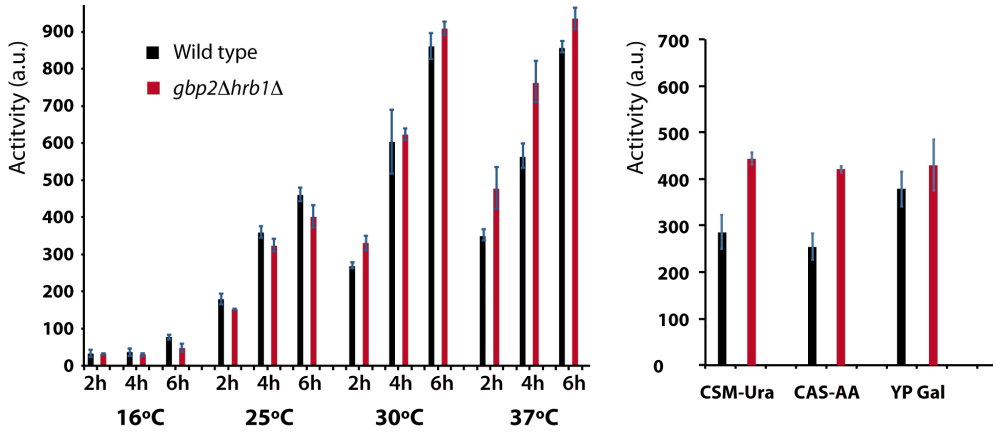


Figure 3.43. – Results of the β-galactosidase assays made under different conditions: changing induction times and temperatures in rich media (left) and using different media at 37ºC (right).

To determine if the effect on β-galactosidase activity is correlated with an increase in mRNA levels, a real time PCR was performed and compared with the β-galactosidase activity test. A similar protocol was used to obtain extracts of the four strains (wild type, *gbp2Δ*, *hrb1Δ* and *gbp2Δhrb1Δ*) transformed with the positive control plasmid. The extracts were split into two for simultaneous β-galactosidase assays and for mRNA quantification. This consisted in a first step of RNA purification, a global retrotranscription reaction that copied mRNA information into DNA and finally, a real time PCR using selective primers for β-galactosidase gene (previously validated in Dr. Séraphin lab). Transcript levels were higher in the double deletion background (Figure 3.44) and correlate with also higher enzymatic activity (*i.e.* protein expression), suggesting that the global effect is related to a more efficiency mRNA biogenesis (transcription, export and/or stability).

Summarizing all the results, the implication of Gbp2p and Hrb1p in splicing function or efficiency seems to be improbable. The slight, but significant, differences observed between mutants indicate a global role in mRNA transcription or export (including intron sequences or not). Additionally these differences suggest that the function of Gbp2p and Hrb1p proteins should involve a general minor repressor activity on cell growth likely due to lower mRNA and therefore, protein levels.

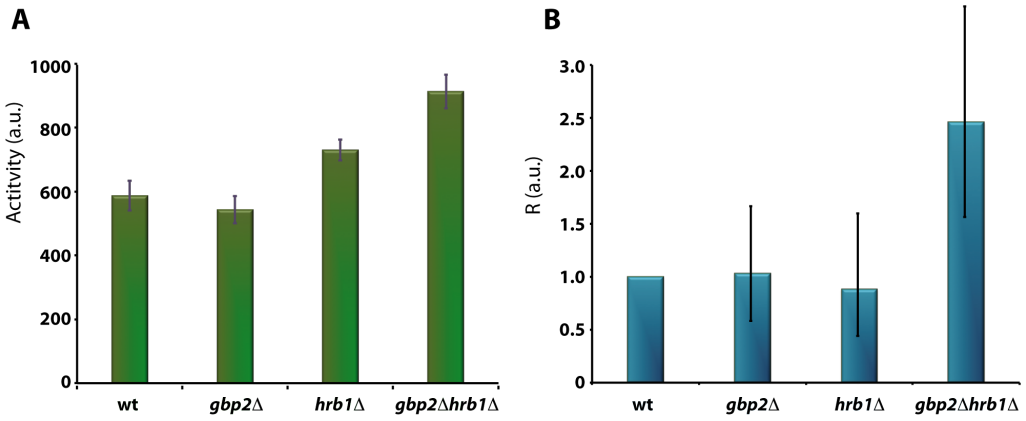


Figure 3.44. – Graphs showing the results of β-galactosidase assays (A) and the quantitative PCR experiment (B) for the different strains tested transformed with positive control plasmid (pBS 169).

3.2.6.3. Tandem affinity purifications and cross-linking studies

Tanden affinity purification (TAP) is an *in vivo* proteomic experiment based on the copurification of several partners that bind to one target protein. This target protein is fussed to the TAP-tag (calmodulin binding peptide + TEV proteolysis site + protein A) in genomic background by typical yeast cloning methods. The purification consists of two affinity steps separated by a TEV digestion and a final analysis by SDS-PAGE identifying the observed bands (the target protein and the copurified ones) by mass spectroscopy peptide footprint (Puig et al., 2001). Previously to this work, there were experimental data available for Gbp2p and Hrb1p binding proteins (including TAP-tags). These data revealed physical and genetic interactions with many proteins, most of which are part of the THO/TREX complex (Hurt et al., 2004). The assays also demonstrated that the interactions are RNA-independent.

The experiments realized in this work were made with the purpose of identifying which of the recognized proteins by Gbp2p (Hrb1p) may interact through its C-terminal. To perform that, some different strains were prepared fussing the TAP-tag in C-terminal variations of *GBP2* sequence. The first assay consisted on a comparison of the copurified proteins by wild type Gbp2p and a version of this protein lacking the C-terminal RRM domain. In addition, two different backgrounds were explored: wild type one and a background lacking Hrb1p protein, obtaining these four combinations: *GBP2-TAP*, *GBP2ΔRRM3-TAP*, *GBP2-TAP+Δhrb1*, *GBP2ΔRRM3-TAP+Δhrb1*.

The results of these TAP experiments are shown in Figure 3.45, where a comparison of Gbp2p-TAP and Gbp2pΔRRM3-TAP copurified proteins shows the disappearance of two characteristic bands in the RRM3 deletion TAPs. These bands were identified by mass spectrometry as Tho2p and Hrp1p proteins, which are elements of THO complex (Strasser et al., 2002). In addition, the presence or absence of Hrb1p in the cell does not alter the protein copurification patterns.

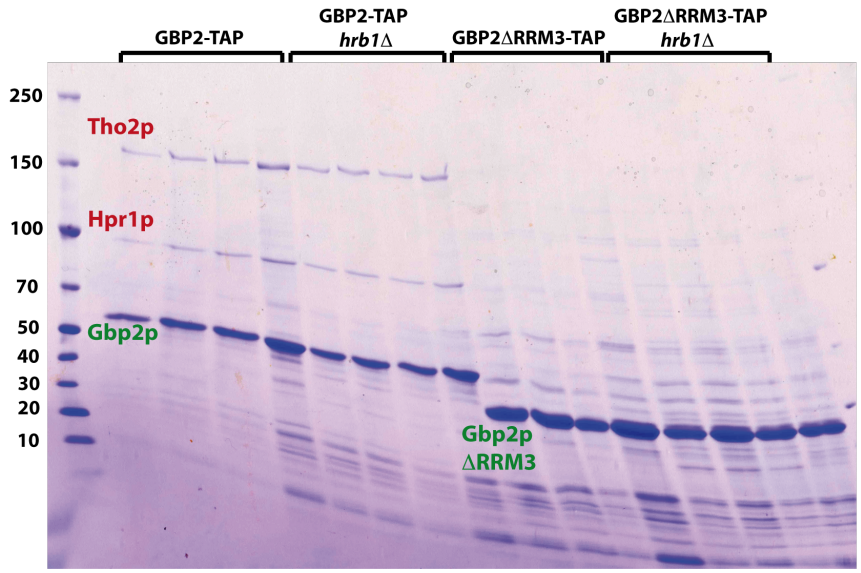


Figure 3.45. – Gradient (5-20%) SDS-PAGE of the eluted fractions of each TAP: lines 1-4 correspond to four GBP2-TAP elutions, 5-8 correspond to GBP2-TAP *hrb1Δ* elutions, 9-12 to GBP2ΔRRM3-TAP elutions and 13-16 correspond to GBP2ΔRRM3-TAP *hrb1Δ* ones.

These results suggest that RRM3 domain helps to recruit the THO complex, although it is a conclusion obtained by a loss of activity. To confirm this hypothesis two other tandem affinity purification experiments were performed: *GBP2ΔRRM12*-TAP (without RRM12 tandem) and the *NPL3-GBP2RRM3*-TAP chimera in which the C-terminal Gbp2p RRM3 domain was fussed to the C-terminal of Npl3p. To confirm that Npl3p is unable to bind the THO by itself, a third *NPL3*-TAP strain was also prepared as negative control (Hurt et al., 2004).

A comparison between the *GBP2ΔRRM12*-TAP and the *GPB2*-TAP shows that the deletion mutant retains the ability to recruit THO (Figure 3.46.A). The Tho2p band (the highest in the gel) is still present in *GBP2ΔRRM12*-TAP, however the stability of this protein construct seemed to be lower than the wild type Gbp2p (by western-blot), and the low levels of Tho2p

do not allowed to identify the protein by mass spectrometry. In addition, these results confirm that nucleic acid binding (that resides on the RRM12) is not required for THO binding. On the other hand, the comparison between *NPL3*-TAP and *NPL3-GBP2RRM3*-TAP (Figure 3.46.B) clearly showed that the Npl3p protein acquires the ability to recruit THO when the RRM3 is fussed to its sequence. In this occasion, the Tho2p band could be confirmed by mass spectrometry (peptide footprinting). These results agree with previous reported data also indicating that Npl3p does not interact with any of the subunits of THO complex (Hurt et al., 2004).

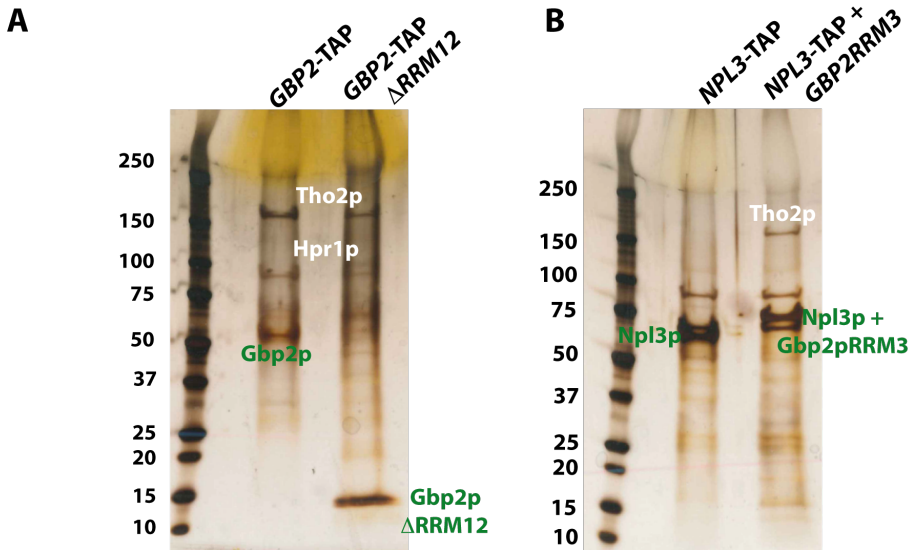


Figure 3.46. – Gradient (5-20%) SDS-PAGE of *GBP2*-TAP and *GBP2ΔRRM12*-TAP elution fractions (A) and *NPL3*-TAP and *NPL3-GBP2RRM3*-TAP elution ones (B).

These tandem affinity purification results assign a function to the RRM3 domain in Gbp2p (Hrb1p): the recruitment of THO complex. Although the protein involved in the direct interaction could not be assigned, the TAP experiments unambiguously identify the RRM3 as the platform that recognises THO complex by the interaction with one or more of the proteins present in this complex.

Finally, cross-linking experiments were performed to find the protein binding partner of the RRM3 domain. THO complex proteins are difficult to manipulate, but recently THO complex could be purified from yeast extract in enough concentration to obtain an electron microscopy low-resolution structure and to map some proteins on it (Peña et al., 2012).

The first attempted experiment consisted in obtaining the purified THO complex using *THO2*- or *HRP1-TAP* tagged strains, and then performing the *in vitro* crosslinking reactions using Gbp2p constructs. Regrettably, the achieved yields in all purification attempts were very low: THO protein bands were only visible under silver staining and also there were present many impurities.

The alternative approach followed uses a more sensible method for protein detection: western blotting. Four different strains were used for this purpose: *THO2-TAP*, *HRP1-TAP*, *MFT1-TAP* and *THP2-TAP*. All the strains have protein A present in their tags to be easily detected using peroxidase anti peroxidase antibodies (PAP). The experiment consisted of a first affinity purification step of the TAP-tags protein extracts (see 2.6.5 section), then heterobifunctional cross-linking reactions, which were performed by activation of recombinant Gbp2p RRM123 protein with a cross-linking agent (sulfo-NHS-LC-diazirine) and finally this protein was mixed with the different yeast extracts and the second UV-activated cross-linking reaction was carried out. As a result, one (or more) of the four THO proteins is expected to be partially cross-linked to Gbp2p. If so, a loss of intensity in the native protein and/or the appearance of a new the adduct band should be detected.

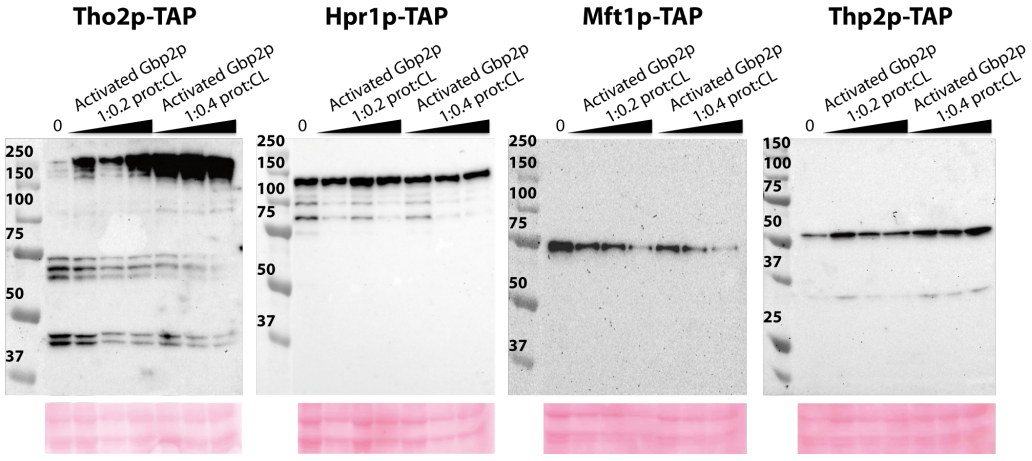


Figure 3.47. –Western blots of the four THO-related protein TAPs after the in vitro cross-linking reactions with Gbp2p RRM123 protein. The reaction control on the first line of each gel shows the extract mixed with buffer and UV-threated as the other lines; the other six show the results for different activated-Gbp2p/TAP mixtures. Three first lanes contain activated protein with a protein : cross-linking agent ratio of 1:0.2 in growing concentrations (1 μM, 5 μM and 10 μM of Gbp2p RRM123) and the next three, activated Gbp2p in a 1:0.4 ratio with the same growing concentrations. Below each gel the corresponding loading control, stained with Ponceau red, is shown.

Many different attempts were performed, trying to optimize several experimental aspects of this complex procedure (lysis methods, crosslinking reaction times, protein concentrations), but no adduct bands were observed in any case. The most relevant result is the stabilization of Tho2p-TAP protein due to Gbp2p RRM123 addition to the media (Figure 3.47).

These results provide an indirect evidence of interaction between Gbp2p and Tho2p protein, but the original aim of identifying a binding partner for Gbp2p remains unsolved. The structure of the THO particle identifies the C-terminal domain of Tho2p as one to the protruding features (Peña et al., 2012) (Figure 3.48). This domain is the only THO region capable to bind to nucleic acids, thus it is tentative to speculate that a hypothetical binding of Gbp2p/Hrb1p proteins nearby would place two different nucleic acids binding domains (C-terminal of Tho2p and RRM12 of Gbp2p) spatially close. This possibility needs to be confirmed in future studies and, if so, the derived implications for the Gbp2p-THO relationship further investigated.

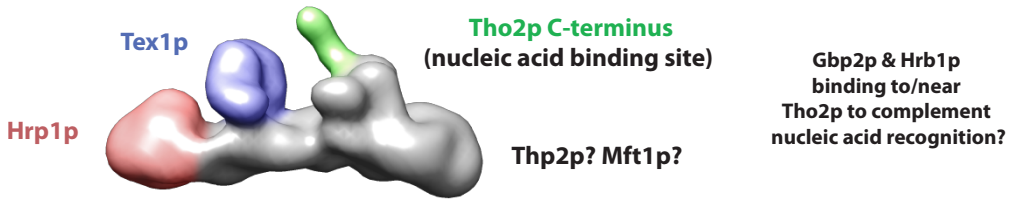


Figure 3.48. – Structure of THO complex (EMD 2053) and the localized subunits on it, extracted from (Peña et al., 2012).

3.3. DISCUSSION

The eukaryotic SR-like protein family is a crucial group of factors participating in mRNA biogenesis. This family is defined by two architectural features: a SR domain made of several arginine and serine dipeptide repeats and at least one RNA recognition motif (RRM) (Busch and Hertel, 2012). The biological function of SR proteins in high eukaryotes (*e.g.* humans) has been related to the recruitment of the splicing machinery to the splice sites, playing a central role in alternative splicing (Fu, 1995; Graveley, 2000). It has been proposed that SR proteins work as splice site enhancers, antagonizing the role of hnRNPs (another broad family of RNA binding proteins). In addition, SR proteins (and hnRNPs) can also have functions in other events such as mRNA export (Huang and Steitz, 2001; Huang et al., 2004), translation regulation (Sanford et al., 2005; Sanford et al., 2004) and genomic stability (Labourier et al.,

1998; Xiao et al., 2007). In summary, the interplay between these two protein families determines the fate of the nascent mRNP particles and their stability.

In yeast, the SR protein group has only three members (Gbp2p, Hrb1p and Npl3p), a small number if compared to the 13 members of the human SR family (SRSF1 to SRSF13). These three proteins are further divided in groups: Npl3p, which share functional and architectural similarity with the mammalian SR proteins and the paralogs Gbp2p and Hrb1p that present different domain architecture and also different roles. The two subfamilies differ in the location of the SR rich region, which is C-terminal to the RRM2 in Npl3p (as in most of the SR proteins) and at the N-terminal part (in their 20 first amino acids) in Gbp2p/Hrb1p. But the biggest difference between these two groups is the presence of an extra third RRM domain in the C-terminus of Gbp2p and Hrb1p with apparently no homology with other RRMs in SR-like proteins. Indeed, some authors do not classify Gbp2p/Hrb1p as SR proteins due to the presence of this extra domain (Manley and Krainer, 2010).

Even more important differences are found regarding their biological function. Npl3p is essential for many processes in mRNA metabolism: transcription elongation and termination (Dermody et al., 2008), splicing (Kress et al., 2008), mRNA export (Lee et al., 1996) and translation initiation control (Windgassen et al., 2004). By contrast, Gbp2p and Hrb1p does not seem to be essential for cell viability and their role in mRNA biogenesis is unknown, appearing as simple passengers in many processes along the pathway (transcription, export, etc). Despite all the differences between Gbp2p/Hrb1p and canonical SR proteins, the results of this thesis reveal that these proteins share crucial structural and probably functional features with human members of the SR family.

3.3.1. Nucleic acid binding of yeast SR proteins

The results of this work for Gbp2p/Hrb1p and previous data for Npl3p (Deka et al., 2008) show that all the RRMs, with the exception of Gbp2p/Hrb1p RRM3 can bind RNA/DNA. The NMR structures of Hrb1p RRM1 and RRM2 domains are very similar to those of Npl3p RRM1 and RRM2 domains (pairwise RMSD values of 1.7Å for RRM1s and 2.2Å for RRM2s). (Figure 3.49). The most relevant differences from the structural point of view are in loops 3 (between β2 and β3). This is longer and His-rich in Hrb1p RRM1, thus its net charge could be easily modulated by pH changes. A heavily positive loop 3 could affect RNA binding in a similar way to that recently proposed for TIA-1 RRM3 (Cruz-Gallardo et al., 2013). Unfortunately the

chemical exchange regime prevents the observation of the signals of these loops and thus their implication in RNA binding could not be investigated. The structures of Gbp2p RRM1 and RRM2 domains were not calculated, however several evidences (CSI, circular dichroism, bioinfomatic analysis) suggest that they must be very similar to Hrb1p, therefore the above discussion can apply to Gbp2p RRM1 and RRM2.

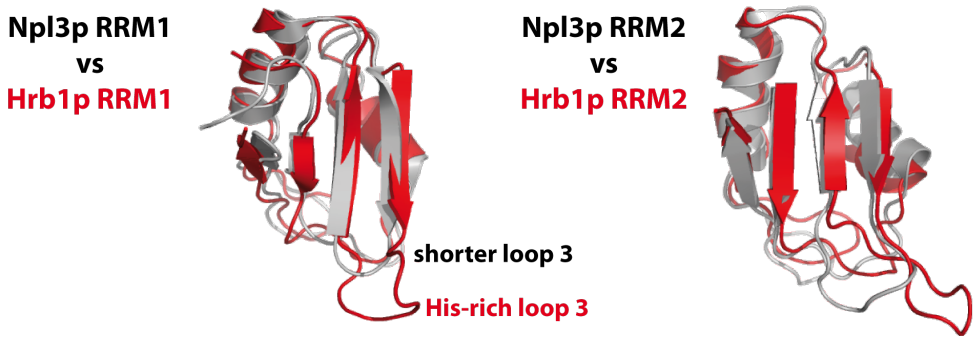


Figure 3.49. – Structural alignment between RRM1s and RRM2s of Hrb1p (this work) and Npl3p (PDBs: 2OSQ and 2OSR).

Npl3p, Gbp2p and Hrb1p share common features on their RNA binding modes. The three RRM1 domains interact with RNA through the classic interface (β-sheet). The binding affinity for Npl3p RRM1 was determined in the μM to mM range. Given that Hrb1p and Gbp2p lack some conserved aromatics on their RNP1/2 motifs (Figure 3. 29), it would be expected a lower affinity for them.

RRM2 binding of all proteins exhibits more interesting traits, presenting a novel RNA binding platform identified very recently (Clery et al., 2013), which involves a conserved sequence in α_1-helixes (Figure 3.37). The new interface is structurally compatible with the β-sheet and binding occurs simultaneously to both in Gbp2p. Indeed the participation of both interfaces might explain why Gbp2p RRM2 domain binds tighter that RRM1, and the corresponding ones in Hrb1p (EMSA in this work). The same was reported for Npl3p (Deka et al., 2008), but not detected in Hrb1p.

Gbp2p, Hrb1p and Npl3p have similar sequence specificity and recognize GU rich sequences (GGU motif) with high selectivity [(Deka et al., 2008) and data of this work], and it is probably commanded by RRM2 domains. The binding mode of these domains can be discussed on the basis of the SRSF1 RRM2 (Clery et al., 2013) (Figure 3.37). The interfacial residues in the α_1-

helix and in the β_2-strand are identical between Npl3p and some human SR proteins. In the case of Gbp2p and Hrb1p, the side chains contacting the guanine pair are kept (-WQ-LKD), whereas the other two residues (S--D--- in Npl3p and human SR proteins and N--(A/S)--- in Gbp2p/Hrb1p) participate in the recognition of the third base, an adenine in Npl3p and probably an uracil in Gbp2p/Hrb1p. The recognition motifs for Hrb1p (KC**YGSU** with K=G/U, Y=C/U, S=C/G) and Gbp2p (H**UGGU**W with H=A/C/U and W=A/U) proteins, identified by SELEX (Riordan et al., 2011), overlap (in bold) in the sequence that contains the RRM2 binding site. Therefore, it could be proposed that this domain (and not RRM1) is the responsible of the RNA specificity of these proteins.

Finally, the RRM1-RRM2 linker is shorter in Npl3p (8 residues) than in Gbp2p/Hrb1p (20 residues), which in principle would favour a stronger tandem effect on the earlier. However biophysical data suggest that Gbp2p/Hrb1p RRM1 and RRM2 are more stable when part of the RRM12 construct and the NMR data (comparison across RRM1, RRM2 and RRM12 construct) is compatible with the presence of RRM1-RRM2 contacts in the RRM123 one. If confirmed, these contacts could help to keep the two RRMs closer, which could benefit RNA recognition.

3.3.1.1. Implications of the novel RNA interface for Gbp2p/Hrb1p phosphorylation

The novel RNA recognition surface discovered in RRM2 domains (called pseudo-RRMs in mammals) also works as binding interface for SR protein kinase family (SRPKs) (Ngo et al., 2008). Many of the α_1-helix amino acids involved in RNA binding interact with a conserved loop (Trp-Gly-His) in the kinase structure (Figure 3.50).

SRPKs hiperphosphorylate the SR proteins, controlling their cellular localization. In the case of Gbp2p, Hrb1p and Npl3p the interaction with their kinase (Sky1p) has been confirmed (except some controversial results with Hrb1p) (Porat et al., 2006), and probably takes place by a similar recognition mechanism that in the human case (key structural elements are conserved in both the SR proteins and the kinase). However the enzymatic mechanism is likely to be different for Gbp2p and Hrb1p. Phosphorylation of human SRSF1 by and SRPK1 proceeds in three steps (Ghosh and Adams, 2011). The first one is this interaction between RRM2 and the so-called the activation loop in the kinase. This allows the SR region (in the C-terminal part of the RRM2) to position in the kinase docking groove for later feed it to the active site to perform the phosphorylation of multiple serine in a processive way. The SR

domain is modified from C to N-terminal direction, until the RRM2 is reached. Unfolding of the β_4-strand of RRM2 and positioning it in the docking groove promotes termination of the process. A similar mechanism could be proposed for Npl3p, because all requirements are conserved; but not for Gbp2p and Hrb1p because their SR regions are at the N-terminal of the protein. Despite this, Gbp2p and Hrb1p could conserve the first SR protein-kinase interacting step and perhaps follow a distributive enzymatic mechanism to modify their SR region.

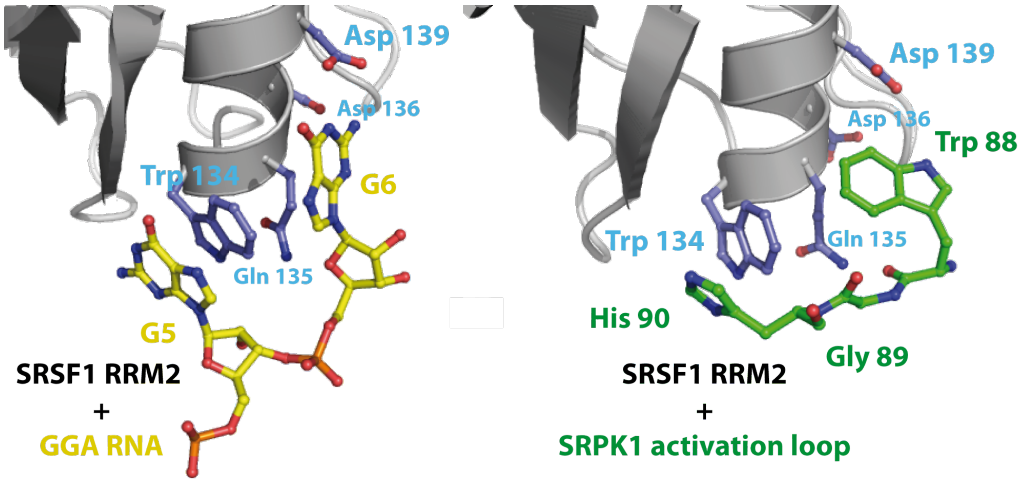


Figure 3.50. - Comparison between the RNA and protein binding modes of pseudo-RRM domain of SRSF1 (RRM2) (PDBs: 2M8D and 3BEG) (Clery et al., 2013).

3.3.2. THO recruitment: the role of the singular RRM3 domains

In comparison with Npl3p and SR proteins, Gbp2p and Hrb1p contain a C-terminal extra RRM , that has a novel N-terminal α_0-helix and no ability to bind RNA/DNA. Instead, its specific function seems to be the recruitment of the THO complex (or recruitment to) (this work), a key step for Gbp2p and Hrb1p mRNA loading. In contrast, Npl3p binds to the nascent mRNA in a THO independent manner (Hurt et al., 2004).

The comparison of fungal homologous (Appendix 3) uncovers a spot of totally conserved residues in α_1-helix (DXXDXX) and loop 5 (YXYX) (Figure 3.51). The lack of conservation in other exposed regions (including the new helix) suggest that this might be the THO interaction site. In support of this hypothesis, a searching of this motif across the 1600 human RRM domains retrieved only SRSF1 and SRSF9 RRM1. The arrangement of these residues in the SRSF1 RRM1 structure (Figure 3.51) shows beyond any doubt that this is truly a

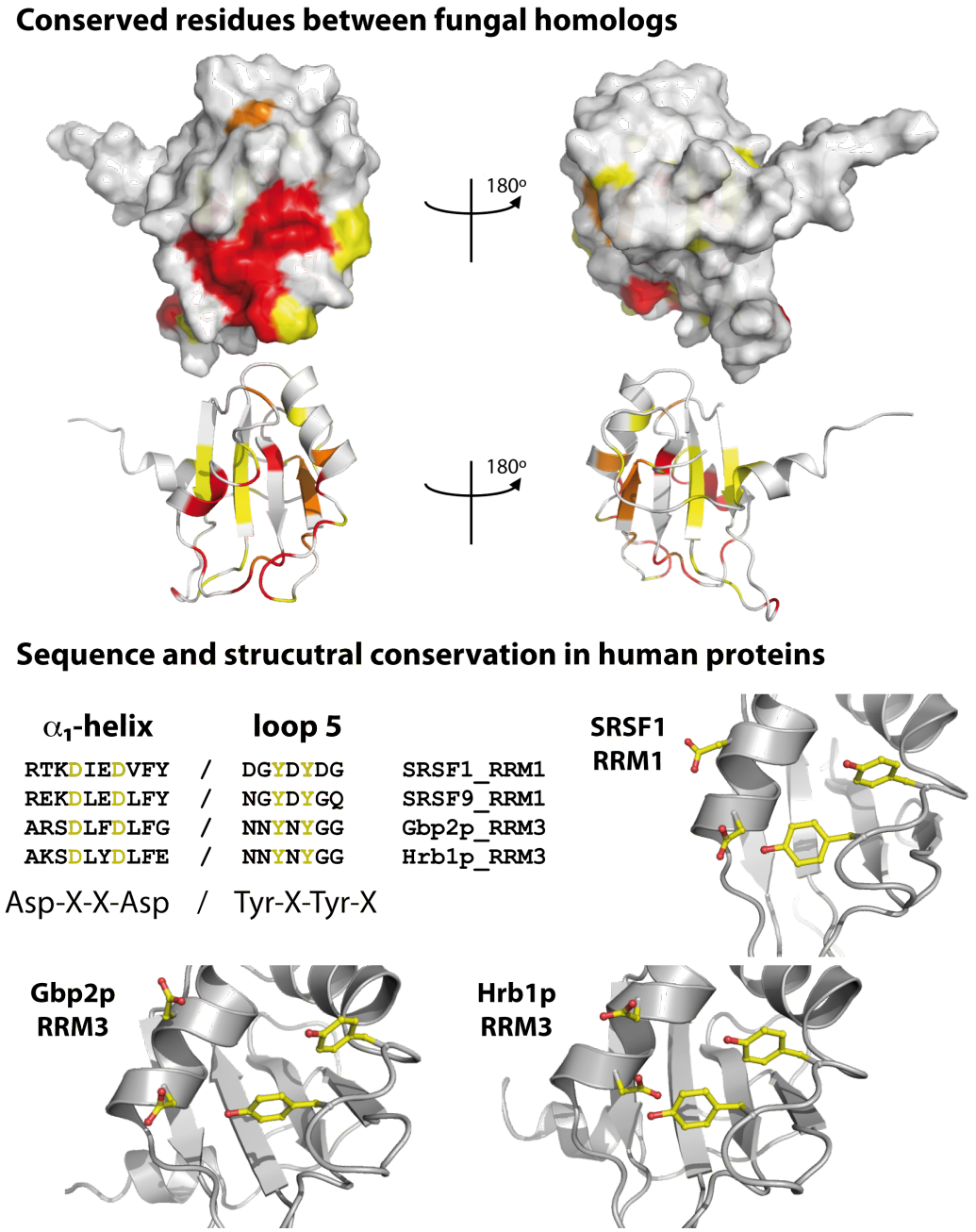


Figure 3.51. – (Up) Conservation of residue positions along different fungal RRM3 domains (from homologs to Gbp2p and Hrb1p). Totally conserved positions in red, almost totally conserved ones in orange and highly conserved in yellow. (Down) Detail of the conserved putative binding surface in SRSF1 RRM1 (PBD: 1X4A), Gbp2p RRM3 and Hrb1p RRM3; and sequence alignment of the studied proteins with SRSF1 and SRSF9.

conserved interface. SRSF1 and 9 are important alternative splicing factors and it is known that human THO complex is loaded to mRNA in a splicing-dependent manner (Masuda et al., 2005). Thus this putative interaction surface in SRSF1 and SRSF9 may play an important role in human THO recruitment.

3.3.3 Are Gbp2p and Hrb1p the real ancestors of SR proteins?

The interplay between SR and hnRNA proteins is crucial for alternative splicing in metazoans (Busch and Hertel, 2012), but there are no clear representatives of these proteins in yeast. Most of the typical mechanisms of alternative splicing found in vertebrates and plants (*i.e.* exon skipping, intron retention, etc.) are absent in yeasts. Thus it could be expected that the systems that regulate these processes in metazoans have not been developed in *S. cerevisiae*. Only two cases of alternative spliced genes have been reported in yeast: the SRC1 and MTR1 genes (Ares Lab Yeast Intron Database: http://compbio.soe.ucsc.edu/yeast_introns.html). The SRC1 pre-mRNA can undertake alternative splicing via alternative 5' splice site selection, which is regulated by ubiquitine-like protein Hub1 (Mishra et al., 2011). But no RNA binding protein has been reported to affect this process to the date.

The current study have found some key structural aspects of the Gbp2p/Hrb1p proteins:

- They bind to the THO complex through the RRM3 and probably by an interface conserved in SRSF1 and SRSF9 RRM1s (major players in alternative splicing in humans).
- They also bind RNAs with the sequence GGU using a new interface that is common to the RRM2 of SRSF1, 4, 5, 6 and 9 and to yeast Npl3p RRM2. This interaction is important for exon selection in humans.
- The same binding site in Gbp2p/Hrb1p RRM2 is predicted to participate in the recruitment of Sky1 (the yeast SR protein kinase).

These evidences support to propose Gbp2p/Hrb1p as the ancestors of SRSF1 and SRSF9. With this hypothesis in mind, it would interesting to test if these yeast proteins might play a role in splice site selection of genes with non-canonical splice sites. Alternative splicing in yeast has a residual impact on cell viability demonstrated by the observation that the *hub1Δ* mutant does not show a detectable phenotype and does not affect the general splicing efficiency (the *Gbp2Δ/Hrb1Δ* shows similar behaviour). It is only when analysing a small proportion of gene expression profiles (only SCR1) when differences can be really accounted for.

The eventual discovery that Gbp2p/Hrb1p affects the splice site selection in yeast would be of outstanding interest, as it would support that the two mechanisms of alternative splicing (ubiquitin-like and SR/hnRNA proteins) appeared long before than the process itself.

3.3.4. Additional possible roles of Gbp2p and Hrb1p

Participation in alternative splicing is one hypothesis of the function of these proteins in yeast, but at the light of the different obtained data, other putative functions could be proposed. For example these proteins could help THO in nucleic acid recognition, as the complex only binds nucleic acids through the C-terminal peptide of Tho2p. Gbp2p and Hrb1p could assist THO binding it with the last RRM3 domain and the DNA or RNA with the RRM tandem. Such putative function might be a sort of helicase activity for the non-coding DNA strand in the transcription bubble or the nascent mRNA, preventing the formation of secondary structures in these sequences (as Gbp2p is able to prevent the formation of superstable structures like G-quadruplexes).

The hypothetical helicase activity might also be important for telomere biochemistry, an area in which Gbp2p was early implicated. The function that it may have is still unclear. It could be that Gbp2p participates in the localization of the transcription machinery near to the telomers or to the nuclear membrane. Or even it could regulate the transcription of some telomeric sequences.

And finally, they were associated to post-transcriptional regulatory structures in the cytoplasm (Gbp2p in stress granules and Hrb1p in P-bodies) but the growth assays show that these proteins do not give growth phenotypes under several stress conditions tested.

3.3.5 Future perspectives

Taking into account all the different collected data and the variety of the proposals for Gbp2p and Hrb1p functions in the cell, many different further experiments could be performed. First, it would be clarifying to go deeper in the knowledge of the nucleic acid binding mode of Gbp2p and Hrb1p proteins, maybe by obtaining a structure of the complex or by further studies on their selectivity and affinity for different motifs (GGU?, GGG?).

Assuming the hypothesis of the role in alternative splicing, one unavoidable experiment is to confirm that in the Gbp2p and Hrb1p deletion mutant strains the alternative splicing is affected in the very few mRNA that present this process.

And finally, the interaction between the RRM3 of the studied proteins and the THO complex should be deeper analysed, identifying the binding partner, confirming the described putative binding platform and performing genetic interaction studies to obtain phenotypes of Gbp2p and Hrb1p, trying to understand the importance of this proteins in mRNA metabolism.

3.4 REFERENCES

Birney, E., Kumar, S., and Krainer, A.R. (1992). A putative homolog of U2AF65 in S. cerevisiae. Nucleic acids research *20*, 4663.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nature reviews Genetics *13*, 770-780.

Buchan, J.R., Muhlrad, D., and Parker, R. (2008). P bodies promote stress granule assembly in Saccharomyces cerevisiae. The Journal of cell biology *183*, 441-455.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. Wiley interdisciplinary reviews RNA *3*, 1-12.

Byrne, K.P., and Wolfe, K.H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome research *15*, 1456-1461.

Chiodi, I., Corioni, M., Giordano, M., Valgardsdottir, R., Ghigna, C., Cobianchi, F., Xu, R.M., Riva, S., and Biamonti, G. (2004). RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic acids research *32*, 4127-4136.

Clery, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. Current opinion in structural biology *18*, 290-298.

Clery, A., Sinha, R., Anczukow, O., Corrionero, A., Moursy, A., Daubner, G.M., Valcarcel, J., Krainer, A.R., and Allain, F.H. (2013). Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. Proc Natl Acad Sci U S A *110*, E2802-2811.

Cruz-Gallardo, I., Aroca, A., Persson, C., Karlsson, B.G., and Diaz-Moreno, I. (2013). RNA Binding of T-cell Intracellular Antigen-1 (TIA-1) C-terminal RNA Recognition Motif Is Modified by pH Conditions. The Journal of biological chemistry *288*, 25986-25994.

Daubner, G.M., Clery, A., and Allain, F.H. (2013). RRM-RNA recognition: NMR or crystallography...and new findings. Current opinion in structural biology *23*, 100-108.

Dauksaite, V., and Akusjarvi, G. (2004). The second RNA-binding domain of the human splicing factor ASF/SF2 is the critical domain controlling adenovirus E1A alternative 5'-splice site selection. The Biochemical journal *381*, 343-350.

Deka, P., Bucheli, M.E., Moore, C., Buratowski, S., and Varani, G. (2008). Structure of the yeast SR protein Npl3 and Interaction with mRNA 3'-end processing signals. Journal of molecular biology *375*, 136-150.

Dermody, J.L., Dreyfuss, J.M., Villen, J., Ogundipe, B., Gygi, S.P., Park, P.J., Ponticelli, A.S., Moore, C.L., Buratowski, S., and Bucheli, M.E. (2008). Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. PLoS One *3*, e3273.

Dujon, B. (2010). Yeast evolutionary genomics. Nature reviews Genetics *11*, 512-524.

Fu, X.D. (1995). The superfamily of arginine/serine-rich splicing factors. RNA *1*, 663-680.

Fujiwara, N., Nakano, M., Kato, S., Yoshihara, D., Ookawara, T., Eguchi, H., Taniguchi, N., and Suzuki, K. (2007). Oxidative modification to cysteine sulfonic acid of Cys111 in human copper-zinc superoxide dismutase. The Journal of biological chemistry *282*, 35933-35944.

Ghosh, G., and Adams, J.A. (2011). Phosphorylation mechanism and structure of serine-arginine protein kinases. The FEBS journal *278*, 587-597.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., *et al.* (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387-391.

Gilbert, W., Siebel, C.W., and Guthrie, C. (2001). Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. RNA *7*, 302-313.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

Hacker, S., and Krebber, H. (2004). Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. The Journal of biological chemistry *279*, 5049-5052.

Hiraga, S., Botsios, S., and Donaldson, A.D. (2008). Histone H3 lysine 56 acetylation by Rtt109 is crucial for chromosome positioning. The Journal of cell biology *183*, 641-651.

Huang, Y., and Steitz, J.A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. Molecular cell *7*, 899-905.

Huang, Y., Yario, T.A., and Steitz, J.A. (2004). A molecular link between SR protein dephosphorylation and mRNA export. Proc Natl Acad Sci U S A *101*, 9666-9670.

Hurt, E., Luo, M.J., Rother, S., Reed, R., and Strasser, K. (2004). Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. Proc Natl Acad Sci U S A *101*, 1858-1862.

Ihara, K., Fujiwara, N., Yamaguchi, Y., Torigoe, H., Wakatsuki, S., Taniguchi, N., and Suzuki, K. (2012). Structural switching of Cu,Zn-superoxide dismutases at loop VI: insights from the crystal structure of 2-mercaptoethanol-modified enzyme. Bioscience reports *32*, 539-548.

Juneau, K., Palm, C., Miranda, M., and Davis, R.W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. Proc Natl Acad Sci U S A *104*, 1522-1527.

Konkel, L.M., Enomoto, S., Chamberlain, E.M., McCune-Zierath, P., Iyadurai, S.J., and Berman, J. (1995). A class of single-stranded telomeric DNA-binding proteins required for Rap1p localization in yeast nuclei. Proc Natl Acad Sci U S A *92*, 5558-5562.

Kress, T.L., Krogan, N.J., and Guthrie, C. (2008). A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Molecular cell *32*, 727-734.

Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009). Circular dichroism and conformational polymorphism of DNA. Nucleic acids research *37*, 1713-1725.

Labourier, E., Rossi, F., Gallouzi, I.E., Allemand, E., Divita, G., and Tazi, J. (1998). Interaction between the N-terminal domain of human DNA topoisomerase I and the arginine-serine domain of its substrate determines phosphorylation of SF2/ASF splicing factor. Nucleic acids research *26*, 2955-2962.

Lee, M.S., Henry, M., and Silver, P.A. (1996). A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes & development *10*, 1233-1246.

Lin, J.J., and Zakian, V.A. (1994). Isolation and characterization of two Saccharomyces cerevisiae genes that encode proteins that bind to (TG1-3)n single strand telomeric DNA in vitro. Nucleic acids research *22*, 4906-4913.

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes & development *24*, 1073-1074.

Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS journal *272*, 2118-2131.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes & development *19*, 1512-1517.

Mishra, S.K., Ammon, T., Popowicz, G.M., Krajewski, M., Nagel, R.J., Ares, M., Jr., Holak, T.A., and Jentsch, S. (2011). Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. Nature *474*, 173-178.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. Nature structural & molecular biology *20*, 127-133.

Ngo, J.C., Giang, K., Chakrabarti, S., Ma, C.T., Huynh, N., Hagopian, J.C., Dorrestein, P.C., Fu, X.D., Adams, J.A., and Ghosh, G. (2008). A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. Molecular cell *29*, 563-576.

Pang, T.L., Wang, C.Y., Hsu, C.L., Chen, M.Y., and Lin, J.J. (2003). Exposure of single-stranded telomeric DNA causes G2/M cell cycle arrest in Saccharomyces cerevisiae. The Journal of biological chemistry *278*, 9318-9321.

Peña, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., Czarnocki-Cieciura, M., Piwowarski, J., Tous, C., Aguilera, A., *et al.* (2012). Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. The EMBO journal *31*, 1605-1616.

Pina, B., Fernandez-Larrea, J., Garcia-Reyero, N., and Idrissi, F.Z. (2003). The different (sur)faces of Rap1p. Molecular genetics and genomics : MGG *268*, 791-798.

Porat, Z., Erez, O., and Kahana, C. (2006). Cellular localization and phosphorylation of Hrb1p is independent of Sky1p. Biochimica et biophysica acta *1763*, 207-213.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Ranjbar, B., and Gill, P. (2009). Circular dichroism techniques: biomolecular and nanostructural analyses- a review. Chemical biology & drug design *74*, 101-120.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Current opinion in cell biology *17*, 269-273.

Riordan, D.P., Herschlag, D., and Brown, P.O. (2011). Identification of RNA recognition elements in the Saccharomyces cerevisiae transcriptome. Nucleic acids research *39*, 1501-1509.

Royer, C.A. (1995). Fluorescence spectroscopy. Methods Mol Biol *40*, 65-89.

Sanford, J.R., Ellis, J.D., Cazalla, D., and Caceres, J.F. (2005). Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. Proc Natl Acad Sci U S A *102*, 15042-15047.

Sanford, J.R., Gray, N.K., Beckmann, K., and Caceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. Genes & development *18*, 755-768.

Senger, B., Simos, G., Bischoff, F.R., Podtelejnikov, A., Mann, M., and Hurt, E. (1998). Mtr10p functions as a nuclear import receptor for the mRNA-binding protein Npl3p. The EMBO journal *17*, 2196-2207.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R*., et al.* (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.

Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Luhrmann, R. (2009). Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. Nature structural & molecular biology *16*, 1237-1243.

Windgassen, M., and Krebber, H. (2003). Identification of Gbp2 as a novel poly(A)+ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. EMBO reports *4*, 278-283.

Windgassen, M., Sturm, D., Cajigas, I.J., Gonzalez, C.I., Seedorf, M., Bastians, H., and Krebber, H. (2004). Yeast shuttling SR proteins Npl3p, Gbp2p, and Hrb1p are part of the translating mRNPs, and Npl3p can function as a translational repressor. Molecular and cellular biology *24*, 10479-10491.

Wishart, D.S., and Sykes, B.D. (1994a). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. Journal of biomolecular NMR *4*, 171-180.

Wishart, D.S., and Sykes, B.D. (1994b). Chemical shifts as a tool for structure determination. Methods in enzymology *239*, 363-392.

Xiao, R., Sun, Y., Ding, J.H., Lin, S., Rose, D.W., Rosenfeld, M.G., Fu, X.D., and Li, X. (2007). Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. Molecular and cellular biology *27*, 5393-5402.

Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191-207.

Zhukov, I., Ejchart, A., and Bierzynski, A. (2008). Structural and motional changes induced in apo-S100A1 protein by the disulfide formation between its Cys 85 residue and beta-mercaptoethanol. Biochemistry *47*, 640-650.

4. RNA binding mode of Nab2p zinc fingers

4.1. INTRODUCTION

This chapter is devoted to the structural studies of Nab2p, another nuclear yeast protein related with mRNP biogenesis (Figure 4.1). In contrast with Gbp2p and Hrb1p, this protein is an essential factor for cell survival (Anderson et al., 1993) and its role is better understood than in the previous case. Nab2p has a double function in the nucleus (Hector et al., 2002); it participates in the control of mRNA stability through the regulation of its poly (A) length in 3' (Eckmann et al., 2011; Soucek et al., 2012) and also it is involved in the mRNPs nucleocytoplasmic transport by interactions with components of the nuclear pore complex (NPC) (Green et al., 2002; Marfatia et al., 2003).

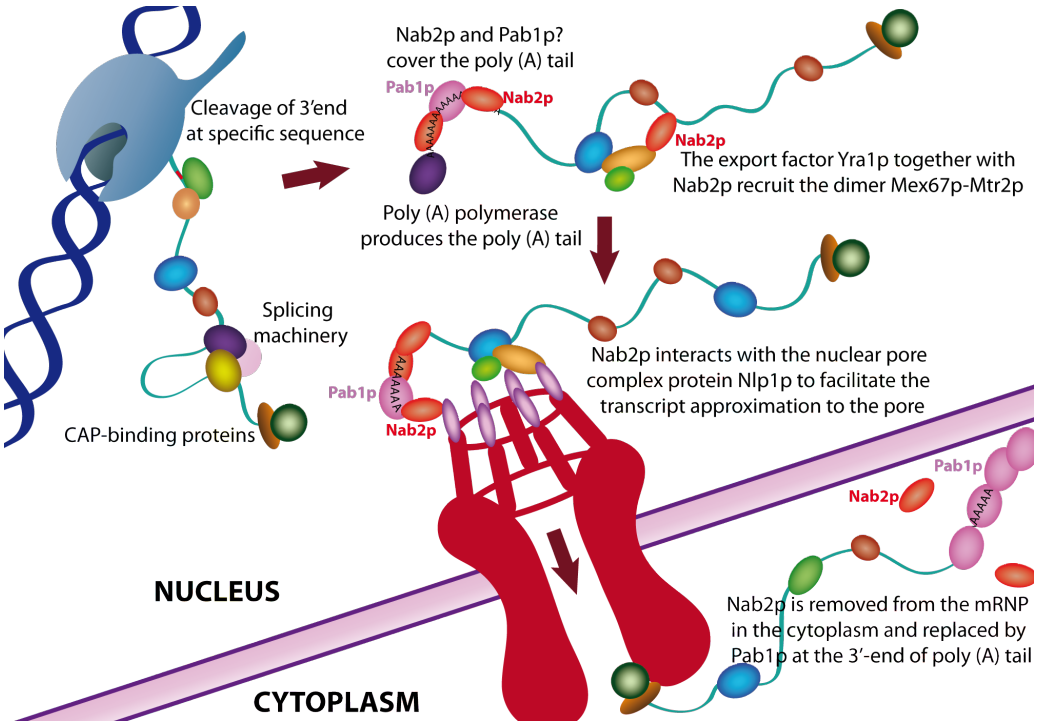


Figure 4.1. – Graphical model of the mRNA 3'-end processing and the nucleocytoplasmic transport in *S. cerevisiae* adapted from (Bergkessel et al., 2009; Soucek et al., 2012).

Transcription termination is triggered by the cleavage of the nascent mRNA at the 3'-UTR, followed by the addition of a poly adenosine RNA tail in the 3'-end (Proudfoot and O'Sullivan, 2002). Nab2p binds this region (Kelly et al., 2010; Kelly et al., 2007; Marfatia et al., 2003) and controls its length, using an unknown mechanism (Kuhn and Wahle, 2004). The size of this tail in yeast ranges between 60 and 80 nucleotides and determines the effectiveness of mRNA

export, mRNA stability and translation in the cytoplasm (Eckmann et al., 2011). Nab2p remains bound to the poly (A) tail during mRNP assembling and participates in the recruitment of Mex67p (Iglesias et al., 2010).

Together with Mex67p and Yra1p, Nab2p recognizes Mlp1p C-terminal domain, a myosine-like protein that is present in the nuclear basket of the NPC, and initiates the docking of the mRNP particle prior to translocation step through the pore (Grant et al., 2008; Green et al., 2003; Vinciguerra et al., 2005). Nab2p is shuttled together with the mRNA to the cytoplasm, where it is dissociated due to mRNP remodelling (Tran et al., 2007). Once released, Nab2p is reimported into the nucleus to participate in a new round of mRNP biogenesis (Marfatia et al., 2003) (Figure 4.1).

Although the two processes in which Nab2p participates could seem to be unrelated, the proper 3′-end formation appears to be the most critical step for the acquisition of export competency and Nab2p, presumably, plays an important role in the coupling of both processes (Niño et al., 2013).

4.1.1. 3'-end processing of mRNAs

This mechanism consists in an endonucleolytic cleavage of the nascent mRNA followed by the poly (A) tail synthesis. The first step is controlled by sequence elements recognition within the untranslated pre-mRNA 3' end (3' UTR), typically four sequence elements (Zhao et al., 1999). This sequences direct the positioning of the processing complex (composed by more than 20 proteins) that performs the RNA cleavage at a specific site [(Niño et al., 2013) and references therein]. Once the transcript has been cleaved, poly (A) polymerase (Pap1p) adds the 60 – 80 adenosines that comprise the poly (A) tail, which is immediately protected by poly (A) RNA binding proteins (PABPs) against the nuclease activity of the nuclear exosome (Amrani et al., 1997; Minvielle-Sebastia et al., 1997).

Three major families of PABPs have been described in eukaryotes: cytoplasmatic PABPC1, that contains four RRM domains displayed in tandems, nuclear PABPN1 with only one RRM and also nuclear Nab2p, which contains seven CCCH-type zinc finger motifs (Kuhn and Wahle, 2004). All of them differ in the mechanisms of polyadenosine RNA recognition. The PABPC1 forms linear deposits on the poly (A) tail in which each protein covers around 12 nucleotides using the first two RRMs (Deo et al., 1999). Both nuclear PABPs bind to the poly (A) tail and to

components of polyadenylation machineries simultaneously (Dheur et al., 2005; Hector et al., 2002; Kuhn et al., 2009). Furthermore, other factors of the nuclear exosome (Lemay et al., 2010; Schmid et al., 2012) are involved in the poly (A) tail length control (Eckmann et al., 2011; Soucek et al., 2012), perhaps competing with these PABPs.

In yeast, Pab1p plays the role of its homolog PABPC1, but there is not a clear homolog of PABPN1. Its most similar protein, Sgn1p, has a function in translation, rather than in poly (A) tail length control (Winstall et al., 2000). In reality, Nab2p seems to be the functional homolog of PABPN1, despite it has a totally different domain architecture (Kuhn and Wahle, 2004).

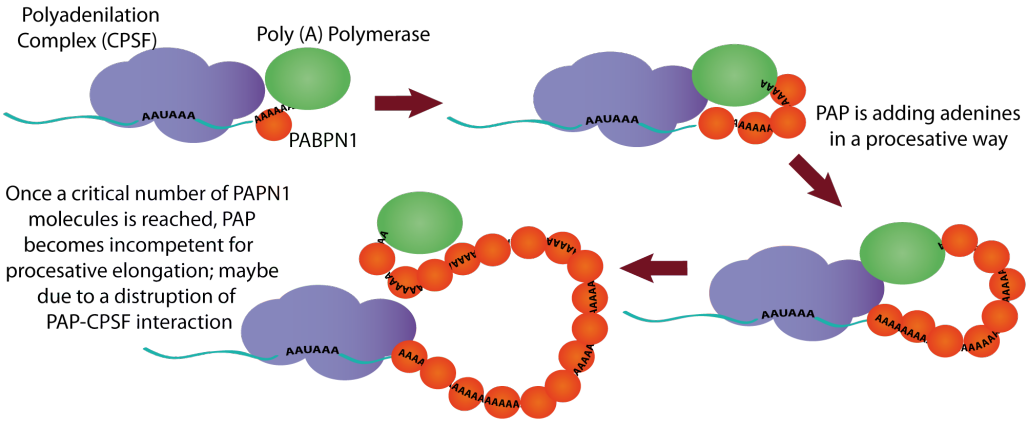


Figure 4.2. – Model of poly (A) tail length control of mammalian poly (A) polymerase by CPSF and PAPBN1, adapted from (Eckmann et al., 2011; Kuhn and Wahle, 2004).

The role of these nuclear PABPs is not only the protection of the 3'-end against nuclease activity but, as it was already pointed, they have essential functions in poly (A) length control and in export. In higher eukaryotes, PABPN1-like protein family regulates the activity of the poly (A) polymerase in conjunction with the Cleavage and Polyadenilation Specificity Factor (CPSF) (Wahle, 1991) (Figure 4.2). This regulation seems to be related with changes in the enzymatic activity of the polymerase that are modulated by the number of PABPN1 copies bound to the poly (A) tail. At early stages, bound PABPN1 proteins stimulate poly (A) polymerase fast processive mode, but when the length raises around 250 nucleotides (the typical poly (A) tail length in mammals), the enzyme switches to slow distributive mode (Kuhn and Wahle, 2004) and dissociates more easily. The details of this mechanism are not known but PAPBN1 plays an essential role in this regulation and some authors propose that the formation of spheric-like PABPN1-poly (A) particles is the crucial event for polymerase

activity control (Eckmann et al., 2011) (Figure 4.2.). In contrast, Nab2p specific mechanism of poly (A) length control is not clear and some evidences also point out to Pab1p (the cytoplasmatic PABP in yeast) as a required partner of Nab2p in this nuclear function (Amrani et al., 1997; Minvielle-Sebastia et al., 1997; Schmid et al., 2012).

4.1.2. Nab2p domain architecture and function

Nab2p is a 525-residue protein that has a multidomain architecture (Figure 4.3). It contains a N-terminal α-helical (PIWI)-like domain that interacts with components of the nuclear pore (Suntharalingam et al., 2004; Zheng et al., 2010), a Q-rich domain with no ascribed function yet, a RGG domain involved in nuclear import (Marfatia et al., 2003) and a C-terminal domain that contains seven CCCH-type zinc fingers, which are responsible of polyadenosine RNA binding (Anderson et al., 1993; Kelly et al., 2010; Kelly et al., 2007). The region that interacts with export factors Mex67p and Yra1p involves both the RGG and the zinc fingers motifs, although the specific binding site has not been determined yet (Niño et al., 2013).

Structural details of the N and C-terminal domains of Nab2p protein are known (Brockmann et al., 2012; Grant et al., 2008; Martinez-Lumbreras et al., 2013). In the zinc finger region, the seven motifs are grouped forming three structurally independent subdomains: two tandems of two zinc fingers each (Zf1-2 and Zf3-4) and the last three zinc fingers region (Zf5-7). Their structures have already been solved by NMR (Figure 4.3). The first tandem Zf1-2 displays a global arrangement in which the two zinc fingers are orientated by the interaction with the C-terminal part of the tandem that folds into a non-canonical α-helix (Martinez-Lumbreras et al., 2013). Zf3-4 tandem folds into a novel arrangement in which the zinc centres define a 2-fold symmetry interaction via the zinc coordinating histidines (Martinez-Lumbreras et al., 2013). And the last three zinc fingers are associated forming another novel structure in which individual fingers are packed together to produce a pseudo-helical arrangement (Figure 4.3) (Brockmann et al., 2012).

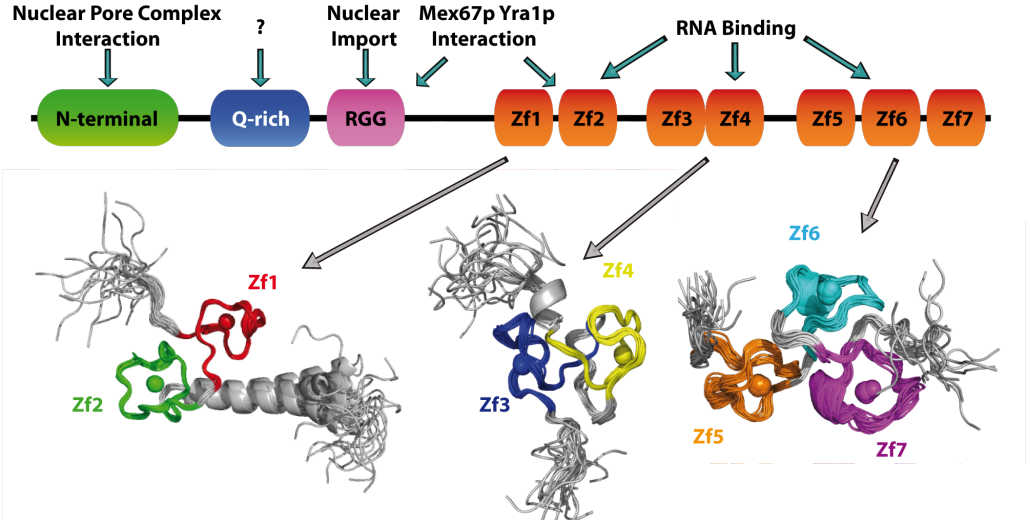


Figure 4.3. – Schematic domain organization of Nab2p protein, indicating the different functions of each element and the structure of zinc finger constructs (PDBs: 3ZJ1 (Zf1-2), 3ZJ2 (Zf3-4) and 2LHN (Zf5-7)).

Nab2p different domains display a variety of interactions with many biomolecules, where the N-terminal part is responsible of protein-protein recognition (proteins in the NPC, Mex67p and Yra1p and nucleus import machinery) and the C-terminal part is the RNA interacting region. This C-terminal part of Nab2p is crucial for yeast survival thus complete or partial deletions of this CCCH domain results in non-viable cells (Marfatia et al., 2003).

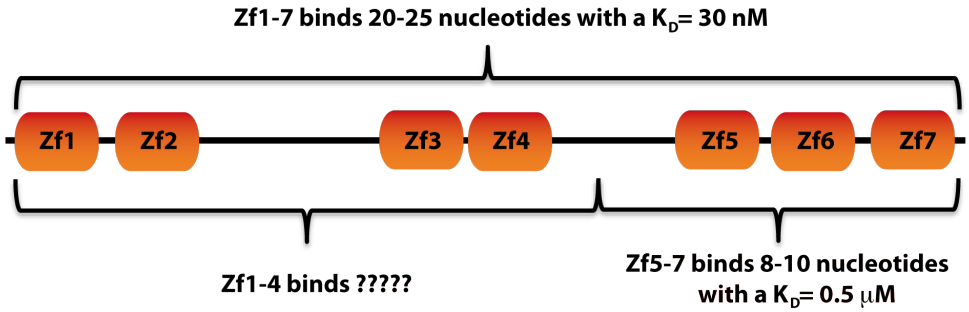


Figure 4.4. – Representation of the Nab2p zinc finger region and the different RNA binding properties of each studied part.

Among the seven zinc finger motifs, the last three (Zf5-7) have been identified as necessary and sufficient for polyadenosine RNA binding (Kelly et al., 2007), although Nab2p ΔZf5-7 mutant still retains some poly (A) RNA binding affinity (Kelly et al., 2010). Full-length protein

displays an affinity 20-fold stronger than Nab2p Zf5-7 construct and the optimal RNA segment covered by the three-zinc finger construct is half the size of the full length Nab2p footprint (Viphakone et al., 2008) (Figure 4.4). All these evidences suggest the existence of other regions that contribute to polyadenosine RNA recognition in Nab2p, a role that has been tentatively attributed to Nab2p Zf1-4.

4.1.3. Objectives

The different studies reported up to now for the RNA binding interaction of Nab2p suggest that the Zf5-7 region is key for RNA recognition, but it cannot reconstitute the whole binding properties of Nab2p full-length protein by itself. Taking this into account, the aim of this study is the characterization of the RNA binding properties of the Zf1-4 region in order to clarify its role, in RNA binding and, if so, to determine its specificity and the optimal length of covered RNA. For that purpose, NMR and fluorescence anisotropy were the tools used in this work.

4.2. RESULTS

4.2.1. Structural analysis of Nab2p zinc finger subdomains

The structures of the RNA binding motifs of Nab2p were already determined and Zf5-7 RNA interaction was also studied (Brockmann et al., 2012; Martinez-Lumbreras et al., 2013). In addition, the RNA binding modes of other CCCH-type tandems like Tis11d (a mammalian protein that binds ARE sequences in mRNA and promotes their deadenilation and subsequent degradation) were also known (Brockmann et al., 2012; Hudson et al., 2004; Kelly et al., 2010; Kelly et al., 2007). Therefore the first performed study consisted on the comparison between the structures of these proteins and those in Nab2p Zf1-4 region focussing on the identification of relevant positions for RNA binding in the two tandem zinc fingers (Zf1-2 and Zf3-4), whose interaction with RNA was not studied yet.

Some residues in Nab2p Zf5-7 and Tis11d are important for RNA interaction and have been postulated as the binding interfaces (Brockmann et al., 2012; Hudson et al., 2004). These residues are located in conserved positions along zinc finger sequence, as it is shown in Figure 4.5. The binding mode in these CCCH type zinc fingers seems to involve a positive charged residue and an aromatic one between the first two coordinated cysteines and another aromatic amino acid between the last coordinated cysteine and the histidine. Such pattern is

observed in some of the zinc fingers located in Zf1-4 region: Zf3 conserves the three positions, Zf1 and Zf2 contain some substitutions and Zf4 is totally degenerated (Figure 4.5). Curiously, the aromatic Tyr position between the second and the third cysteines in Tis11d is not conserved in any Nab2p zinc fingers (neither in the Zf1-4 or in Zf5-7) and in most cases it is replaced by an asparagine residue.

Figure 4.5. – Alignment of the seven CCCH type zinc fingers in Nab2p and Tis11d. The residues involved in RNA binding in Nab2p Zf5-7 and Tis11d and those conserved positions in Nab2p Zf1-4 are highlighted in red and the aromatic residue position in Tis11d also important for RNA interaction, but not conserved in Nab2p (Asn position) is highlighted in green.

It is important to check if conserved key residues for RNA binding occupy equivalent tridimensional locations at the hypothetical interface in Nab2p Zf1-2/Zf3-4 tandems. To analyse this issue, a set of structural alignments between different Nab2p tandems was made. Average RMSD values across the pairwise fit between all combinations of Zf1-2, Zf3-4 tandems and the three different dispositions of the Zf5-7 (Zf56, Zf67 and Zf57) were calculated (Table 4.1).

	Nab2p Zf1-Zf2 pair	Nab2p Zf3-Zf4 pair	Nab2p Zf5-7 Zf5-Zf6 pair	Nab2p Zf5-7 Zf5-Zf7 pair	Nab2p Zf5-7 Zf6-Zf7 pair
Nab2p Zf1-2	-	5.8 ± 0.1	4.2 ± 0.1	6.1 ± 0.1	1.8 ± 0.1
Nab2p Zf3-4	5.8 ± 0.1	-	5.8 ± 0.2	7.4 ± 0.1	6.8 ± 0.1

Table 4.1. – Structural alignments between Nab2p Zf1-2, Zf3-4 and Zf5-7, each value indicates the corresponding average RMSD (± SD) across the pairwise fit between structures (20x20). Backbone atoms, zinc coordinating Cys and His atoms and Zn^{2+} were used for the fitting (120 atoms for each conformer).

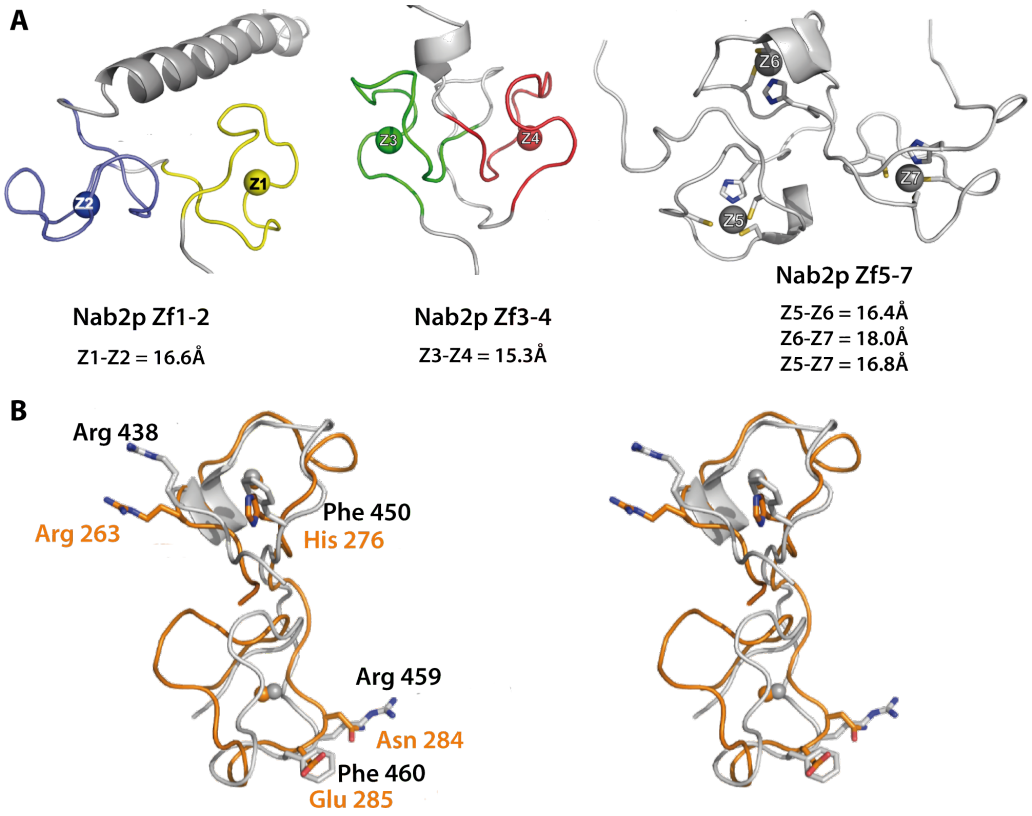


Figure 4.6. – A) Representative arrangements of Nab2p CCCH-type zinc finger structures, indicating the distance between metal ions. B) Stereoview of the structural alignment between Nab2p Zf1-2 (orange) and Zf6-7 (grey), showing the side chain residues involved in RNA binding of Zf6-7 and the equivalent positions in Zf1-2.

The results showed that Nab2p Zf3-4 tandem displays poor similarities with Nab2p Zf1-2 and with the three sub-pair in the structure of Nab2p Zf5-7. This is because the interaction between the two zinc centres defines a shorter distance between zinc ions than in other CCCH-containing structures, as it is shown in Figure 4.6.A. In contrast, the global orientation of Zf1-2 is very similar to the structure of the zinc finger pair 6-7 within Zf5-7 construct, with a mean pairwise RMSD value of 1.8 Å. In Nab2p Zf5-7, the contacts between each pair of zinc fingers provide structural coherence and, similarly, zinc fingers in Zf1-2 tandem also forms a coherent structure, where the C-terminal helix (rather than a third zinc finger as in Zf5-7), plays the structural role to maintain such coherence. The structural alignment (Figure 4.6.B) clearly shows that the positions involved in Zf5-7 RNA binding are not conserved in Zf1-2: among Arg 438, Phe 450, Arg 459 and Phe 460 residues in Zf5-7; only Arg 263 (equivalent to Arg 438) and the aromatic character at His 276 (Phe 450) are conserved in Zf1.

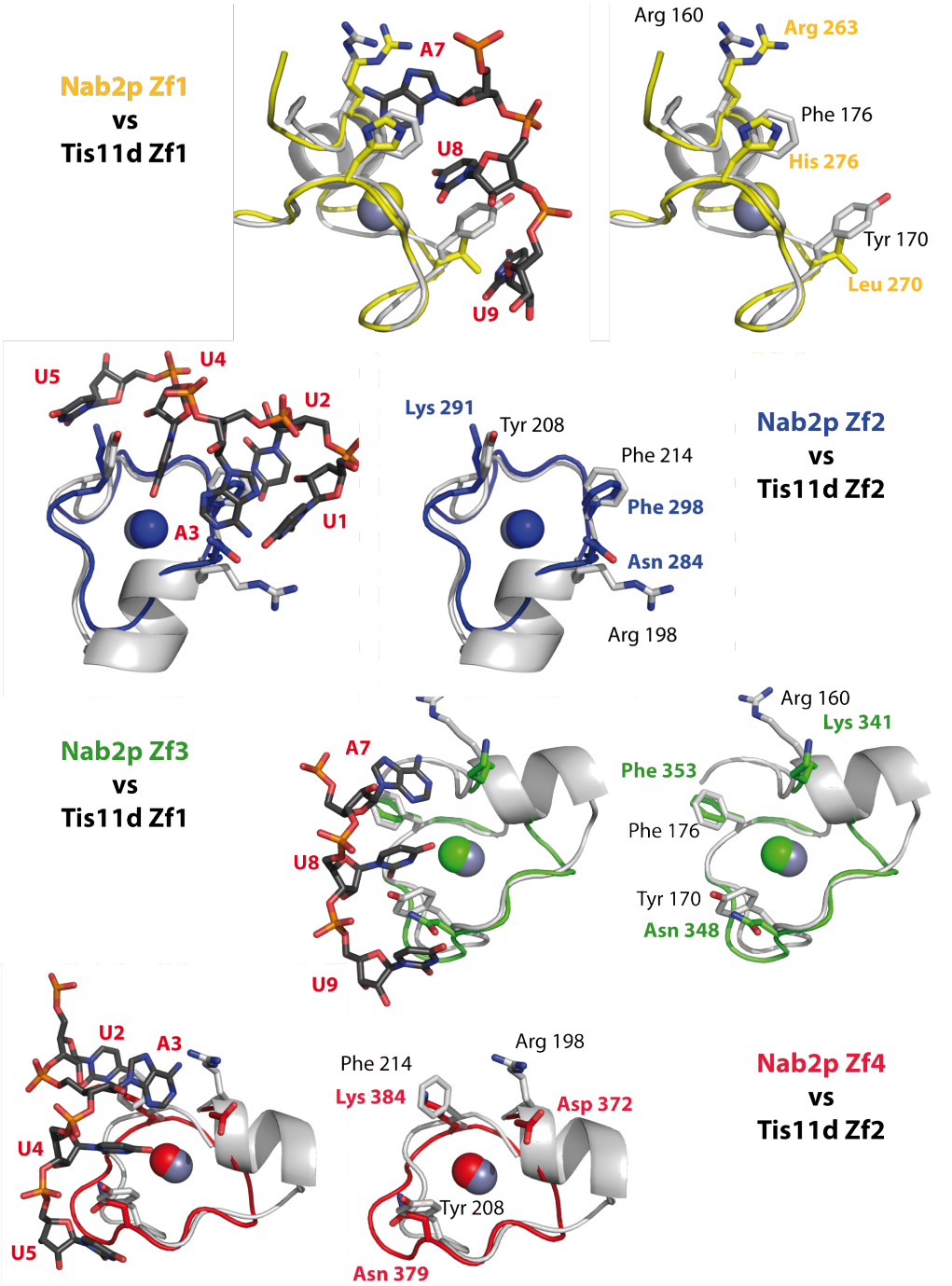


Figure 4.7. – Structural alignments of the single zinc finger domains of Nab2p Zf1-4 (PDB 3ZJ1 and 3ZJ2) with those of Tis11d:RNA complex (PDB 1RGO) (Hudson et al., 2004; Martinez-Lumbreras et al., 2013).

Additional structural comparisons were made with the zinc finger domain of Tis11d, whose protein-RNA (AU rich) complex structure was already determined (Hudson et al., 2004). In this case, a single zinc finger structural alignment was performed due to the different Zn-Zn distance in this complex (25.5 Å) compared with Nab2p structures. In the alignment (Figure 4.7) the residues of Tis11d involved in RNA binding are compared with the same positions in Nab2p zinc fingers. The results of this comparison show that in Zf1-2, Arg263, His 276 (Zf1) and Phe 298 (Zf2) conserve the positive and aromatic character, important for RNA binding in Tis11d (Arg 160, Phe 170 and Phe 214, together with other residues, are involved in RNA bases recognition). In the case of Zf3-4, the conserved residues appear in Zf3 (Lys 341 and Phe 352 equivalent to Arg 160 and Phe 170 in Tis11d) but in Zf4 there is not such conservation. Residues Tyr 176 and Tyr 208 in Tis11d are important for RNA base recognition, but as mentioned above, no aromatic residues are found at this position for any of Nab2p zinc fingers, rather an asparagine residue was found for most of them. These residues are proved to be perturbed upon RNA titration in Zf5-7 (Brockmann et al., 2012), although its role in RNA binding was not explored by mutagenesis.

Summing up all this information, the detected structural and sequence differences in Zf1-2 and Zf3-4 tandems suggest a different RNA binding character of the N-terminal part of Nab2p zinc finger region, that is further studied in the following experiments.

4.2.2. Study of polyadenosine RNA recognition by Zf1-4 construct

Full-length Nab2p binds polyadenosine RNA with high affinity and specificity (K_D ~20-30 nM to A_{25}) (Hector et al., 2002; Kelly et al., 2010; Kelly et al., 2007; Tran et al., 2007), while the Nab2p Zf5-7 construct has been reported to bind 8 to 10 nucleotides with a K_D of 0.5 to 0.1 µM (Brockmann et al., 2012). To complement these studies, the RNA binding mode of Nab2p Zf1-4 construct has been examined by fluorescence anisotropy and NMR.

First, K_D values obtained from fluorescence anisotropy titrations between Nab2p Zf1-4 construct and two fluorescent probes A_{10} and A_{12} were of 4.4 ± 0.4 µM and 2.3 ± 0.2 µM, respectively (Figure 4.8.A and Table 4.2). These data were validated and a possible effect of fluorophore ruled out by the results obtained in competition experiments against the labelled A_{12} probe with unlabelled RNAs of different lengths (8, 10, 12, 14 and 20 nucleotides). They also showed that the affinity reaches a plateau around twelve nucleotides, which defines the optimal RNA length bound to Zf1-4 construct (Figure 4.8.B and C and Table 4.2).

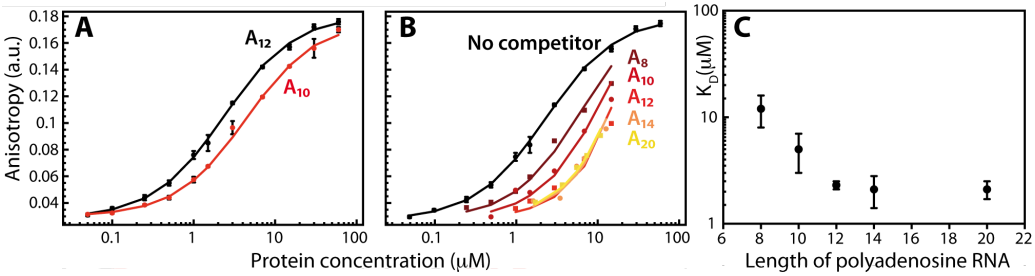


Figure 4.8. – A) Fluorescence anisotropy binding titrations of Nab2p Zf1-4 construct with A_{12} and A_{10} fluorescein labelled RNAs. B) Binding titration curves of labelled A_{12} with Zf1-4 obtained in the presence of 20 μM of various polyadenosine RNAs (unlabelled) as competitors. C) Dependence of the obtained K_D values from fittings in B) upong the length of polyadenosine RNA.

In order to test the specificity of Nab2p Zf1-4 construct, titrations with labelled U_{12} probe and competition curves against A_{12} (with unlabelled U_{12} and $(AU)_6$ RNAs) were performed. The results suggest that Nab2p Zf1-4 displays a modest polyadenosine specificity as binding to U_{12} and $(AU)_6$ is between 3- and 5-fold weaker than to A_{12} (Figure 4.9.A and B and Table 4.2).

Finally, the binding affinities of Zf1-2 and Zf3-4 tandems were explored. Zf3-4 did not produce changes in the anisotropy of the A_{12} probe, while Zf1-2 titration resulted in a weaker binding than the Zf1-4 construct (Figure 4.9.C and Table 4.2). Therefore the binding mode of Nab2p Zf1-4 region must involve both tandems simultaneously.

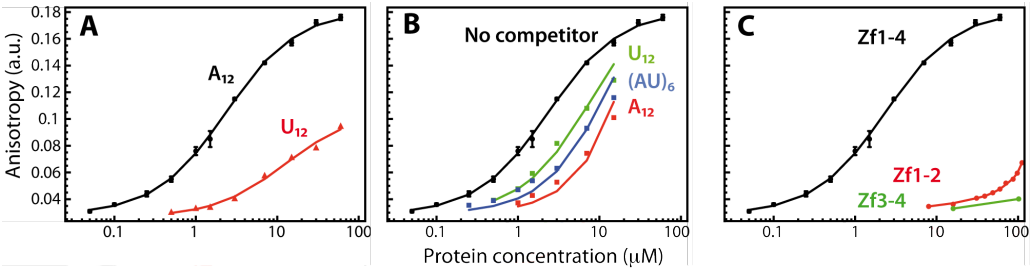


Figure 4.9. – A) Fluorescence anisotropy binding titrations measurements of Nab2p Zf1-4 construct with A_{12} and U_{12} fluorescein labelled RNAs. B) Competition binding titration curves of labelled A_{12} with Nab2p Zf1-4 protein obtained in the presence of 20 μM of A_{12}, U_{12} and $(AU)_6$ unlabelled RNAs as competitors C) Binding fluorescence anisotropy titrations using labelled A_{12} RNA and three different protein constructs: Nab2p Zf1-4, Zf1-2 and Zf3-4.

Protein	Fluorescent RNA	Competitor RNA	ΔG (kcal/mol)	K_D (μM)
Nab2p Zf1-4	f-A_{12}	-	7.56 ± 0.05	2.3 ± 0.2
	f-A_{12}	A_{12}	7.7 ± 0.3	2 ± 1
Nab2p Zf1-4	f-A_{10}	-	7.18 ± 0.05	4.4 ± 0.4
	f-A_{12}	A_{10}	7.1 ± 0.2	5 ± 2
Nab2p Zf1-4	f-A_{12}	A_{20}	7.6 ± 0.1	2.1 ± 0.4
Nab2p Zf1-4	f-A_{12}	A_{14}	7.6 ± 0.2	2.1 ± 0.7
Nab2p Zf1-4	f-A_{12}	A_8	6.6 ± 0.2	12 ± 4
Nab2p Zf1-4	f-U_{12}	-	6.6 ± 0.2	12 ± 4
	f-A_{12}	U_{12}	6.7 ± 0.2	10 ± 3
Nab2p Zf1-4	f-A_{12}	$(AU)_6$	7.2 ± 0.2	7 ± 4
Nab2p Zf1-2	f-A_{12}	-	< 5.0	>190
Nab2p Zf3-4	f-A_{12}	-	-	-

Table 4.2. – Energetic parameters determined by fluorescence anisotropy of Nab2p–RNA interactions using different RNAs and protein constructs.

Apart from fluorescence anisotropy titrations, Nab2p Zf1-4 binding to A_{12} RNA was also studied by NMR. The ^{1}H-^{15}N HSQC spectrum of this protein construct experienced important line broadening upon titration with A_{12} probe (Figure 4.10.A). The exchange regime is heavily dependent on the spectrometer field strength and most of the signals disappear at sub-stoichiometric conditions at 800 MHz (proton field). The quality of the spectrum improves when reaching 1:1 RNA:protein ratio and the spectrum remains invariable at 2:1 RNA excess, suggesting that the stoichiometry of the complex is equimolecular (Figure 4.10.B).

Assignment of the A_{12} RNA bound form of Nab2p Zf1-4 was attempted by comparison with free spectra and using some tridimensional experiments (3D-HNCA and 3D-CBCA(CO)NH), but it could not be fully completed due to the line broadening. Only a 34% of NH signals in Zf1-2 tandem and 72% in the Zf3-4 one were assigned (Figure 4.11.A). The comparison of the chemical shift values between free and bound forms (using the described formula in Equation 2.5. for averaged chemical shift differences) showed that both tandems were perturbed (Figure 4.11.B). However the incomplete map of chemical shift perturbations impedes to get unambiguous conclusions about the RNA interface of Nab2p Zf1-4.

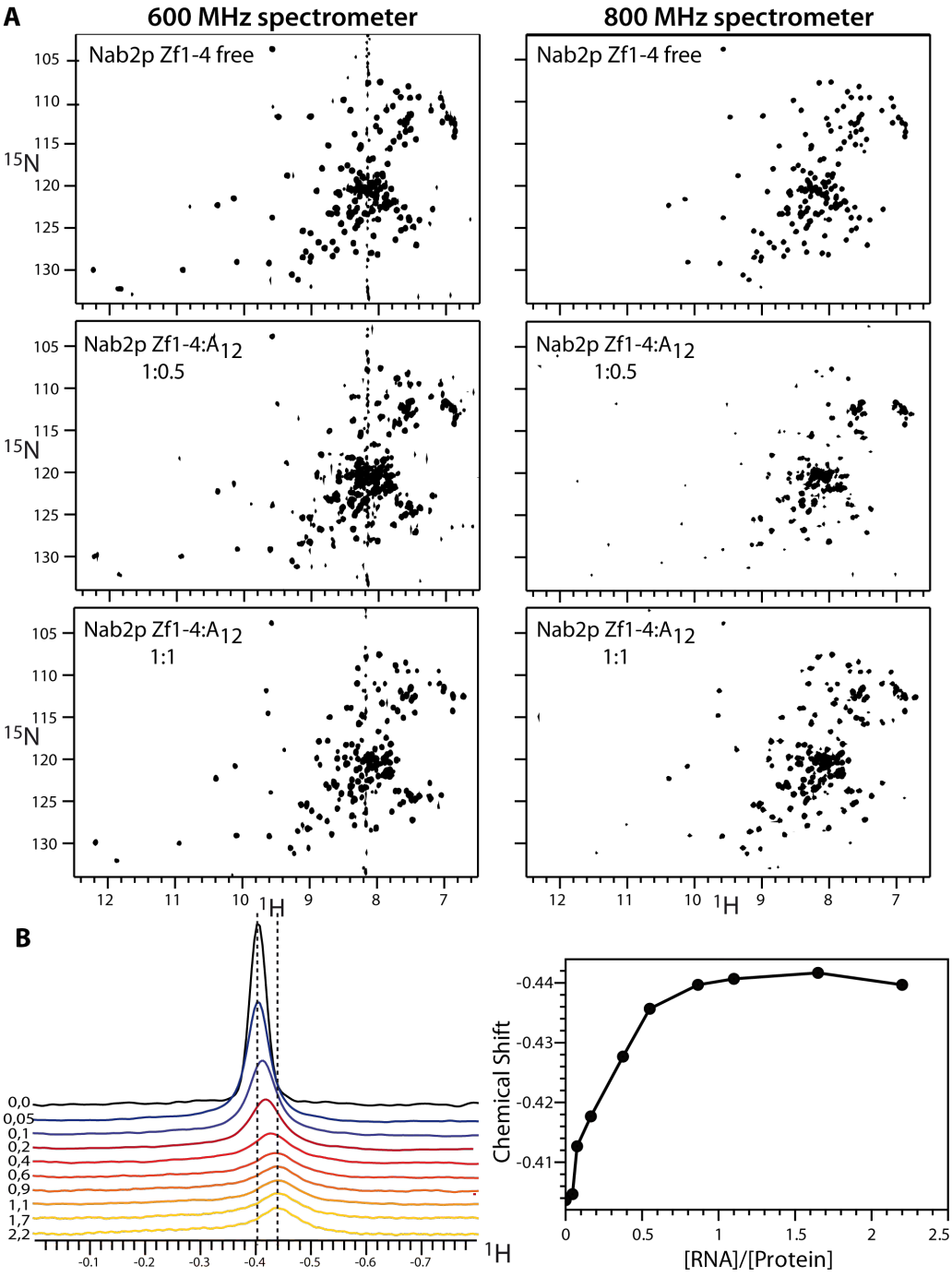


Figure 4.10. – A) NMR titrations of ^{15}N-labelled Nab2p Zf1-4 protein with A_{12} RNA followed by ^{1}H-^{15}N HSQC acquisition at three protein:RNA ratios and in two spectrometers (600 MHz and 800 MHz). B) 1D spectra in methyl region of different titration points of Zf1-4 with A_{12} RNA.

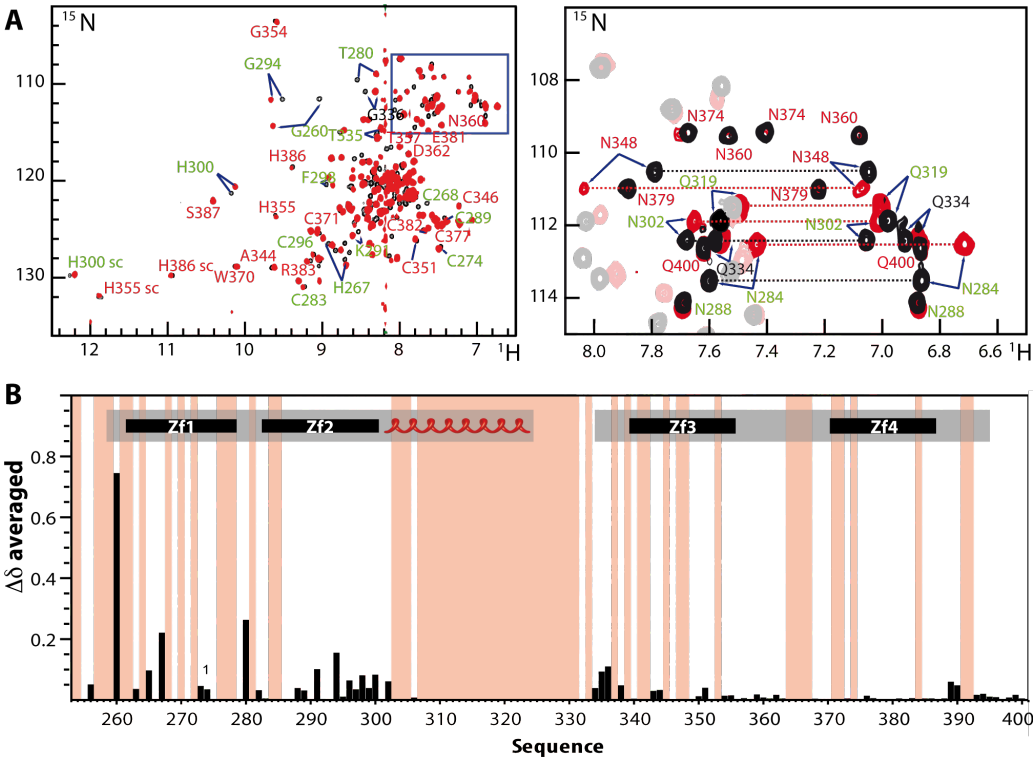


Figure 4.11. – A) Partially assigned ^{1}H-^{15}N HSQC spectra of Nab2p Zf1-4 in free (black) and bound (red) forms B) Histogram representing the averaged chemical shift differences between free and bound (1:1) NH signals of Zf1-4 upon titration with A_{12} RNA. Orange bars highlight the unassigned regions.

Although a clear complete interaction interface could not be derived from the perturbation map, some interesting information about binding could be extracted. Thus, most of the residues in the α-helix of the Zf1-2 region disappeared upon titration, suggesting that this part may participate in RNA binding. In addition, some assigned peaks of the zinc fingers were perturbed (like Asn 284 and Gln 319 side chains or Gly 260, His 267, Thr 280 or Gly 294 NH correlations, highlighted in green in Figure 4.11.A). In the case of Zf3-4 region each zinc finger showed a significantly different behaviour. The side chain of Asn 348 of Zf3 is perturbed upon titration. The structurally equivalent residues Asn 423 and Asn 466 in Nab2p Zf5-7 were also perturbed upon RNA binding (Brockmann et al., 2012) and their structural equivalents in Tis11d: Tyr 176 and Tyr 208 in Tis11d, participate in RNA binding (Hudson et al., 2004). However the equivalent residue in Zf4 (Asn 379) is not perturbed suggesting that Zf4 does not play an active role in RNA recognition. By contrast, some side chain resonances in the C-terminal α-helix of Zf1-2 are perturbed.

Overall, the NMR data suggest a clear participation of Zf1-2 region in RNA binding (including the α-helix) and a possible role of Zf3 but not of Zf4.

4.2.3. Analysis of RNA binding of different Nab2p Zf1-4 mutants

As the RNA binding face of Nab2p Zf1-4 construct could not be clearly determined from the NMR titration experiment, the identification of residues important for RNA binding was addressed throughout the analysis of the affinity of different Zf1-4 mutants.

The first round of Nab2p Zf1-4 mutants were designed on the basis of the individual zinc fingers sequence comparison in Zf1-4 and Zf5-7 constructs (Figure 4.5), structural homology to equivalent protein-RNA complexes (Figures 4.6 and 4.7) and previous mutagenesis data on Nab2p Zf5-7 (Brockmann et al., 2012). Fluorescently-labelled RNA (A_{12}) was titrated with each Nab2p Zf1-4 mutant in an equivalent experimental setup to wild-type construct (Figure 4.12).

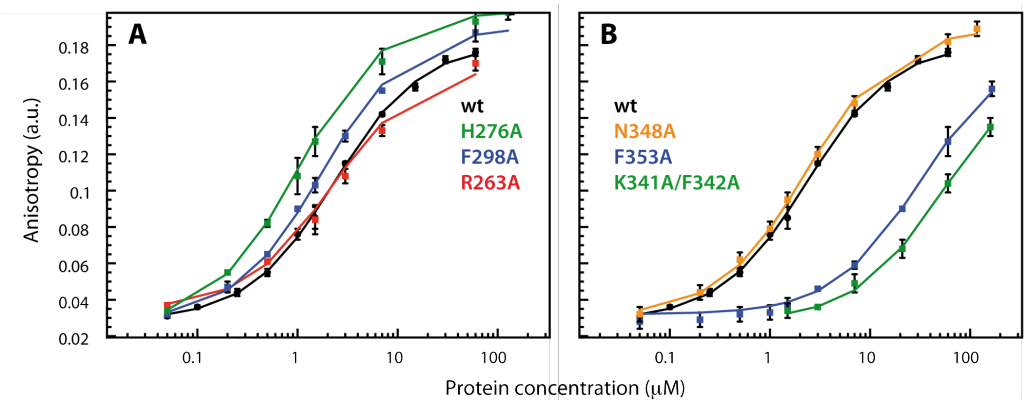


Figure 4.12. – Fluorescence anisotropy binding titrations of fluorescein labelled A_{12} with different mutants of Nab2p Zf1-4 construct: in Zf1-2 (A) and in Zf3-4 tandems (B).

According to these titrations (Figure 4.12 and Table 4.3), alanine substitution in the zinc fingers 1 and 2 at equivalent positions to those important in Zf5-7 (Arg 263, His 276 (Zf1) and Phe 298 (Zf2) similar to Arg 438, Phe 450 (Zf6) and Phe 471 (Zf7)) (Figures 4.5, 4.6 and 4.7) did not cause a significant variation in RNA affinity. This indicates that they are not required for RNA interaction (Figure 4.12.A). Mutants at Arg 438, Phe 450 and Phe 460/Phe 471 in Nab2p Zf5-7 have been reported to decrease the affinity for A_8 between 76- and 148-fold (Brockmann et al., 2012). In contrast, mutations in similar positions in Zf3 (Lys 341/Phe 342

and Phe 353) showed a significant drop in binding affinity (16- and 25-fold; Figure 4.12.B and Table 4.3), pointing that the participation of this zinc finger in the RNA binding follows a similar recognition mechanism than in Nab2p Zf5-7 and other CCCH-protein/RNA complexes (Figures 4.6 and 4.7). This recognition mode proposes Asn 348, a residue conserved in Zf3, Zf4, Zf5 and Zf7 of Nab2p (Figure 4.5), as an important residue in RNA recognition. Indeed, the Asn 348 side chain signals were strongly perturbed in the NMR titrations. However, the alanine mutant showed no variation in the affinity. Surprisingly, mutants in the first tandem seem to alter the mobility of the fluorescent probe, as they presented subtle variations of the anisotropy value of the bound form (Figure 4.12.A).

The above experiments do not reveal which is the RNA recognition mechanism of the Zf1-2 tandem but demonstrate that it should be different to the binding modes in Nab2p Zf5-7 and other CCCH-type protein-RNA complexes (Brockmann et al., 2012; Hudson et al., 2004). As the C-terminal helix signals were strongly perturbed upon NMR titrations and this helix is clearly charged (it contains 8 negative charged residues and 9 positive charged ones) two more mutants were designed changing positive charged residues (conserved and solvent exposed) in the helix to alanine amino acids. These mutants present a decrease in RNA affinity, pointing to an electrostatic interaction between the helix and the RNA (Figure 4.13.A and Table 4.3) as the RNA recognition mechanism of this tandem.

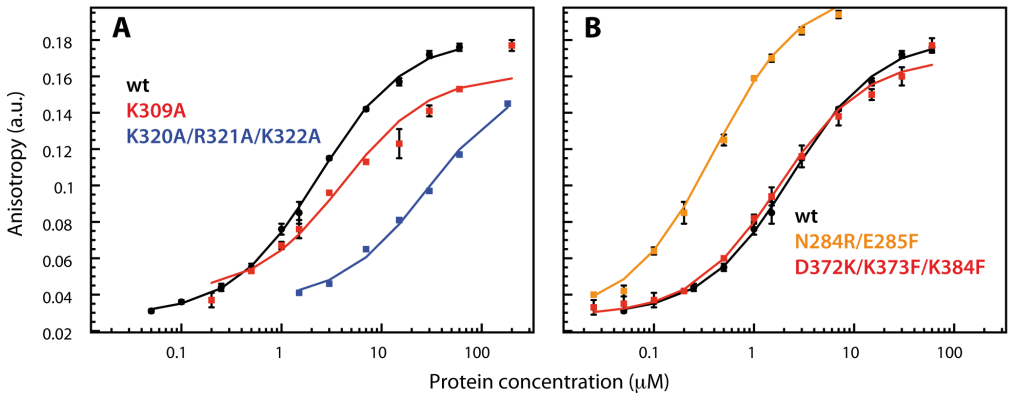


Figure 4.13. – Fluorescence anisotropy binding titrations of fluorescein labelled A_{12} with mutants of Nab2p Zf1-4 construct in the C-terminal helix of Zf1-2 (A) and some other mutations to restore the RNA binding configuration of Zf2 and Zf4 (B).

To get further detail onto the Nab2p Zf1-4 RNA binding mode, a third generation of mutants were designed, seeking to optimize the "classic" binding positions in the Zf1-2 and Zf3-4 tandems. For the N284R / E285F mutant, that restores a configuration of the Zf1-2 tandem similar to Zf6-7 (Figure 4.6), this approach was successful as it showed higher affinity than wild type variant; indeed the K_D value is nearly identical to that of Nab2p Zf5-7 for A_8. In contrast, when Zf4 is mutated in order to reconstitute the typical RNA binding configuration (mutant D372K / K373F / K348F), the affinity practically remained as in the wild type (Figure 4.13.B and Table 4.3).

Protein	Fluorescent RNA	ΔG (kcal/mol)	K_D (μM)
Nab2p Zf1-4 R263A	f-A_{12}	7.6 ± 0.1	2.1 ± 0.4
Nab2p Zf1-4 H276A	f-A_{12}	8.0 ± 0.2	1.0 ± 0.4
Nab2p Zf1-4 F298A	f-A_{12}	7.7 ± 0.1	1.8 ± 0.3
Nab2p Zf1-4 N348A	f-A_{12}	7.6 ± 0.1	2.1 ± 0.4
Nab2p Zf1-4 K341A/F342A	f-A_{12}	5.7 ± 0.2	56 ± 19
Nab2p Zf1-4 F353A	f-A_{12}	6.0 ± 0.1	33 ± 6
Nab2p Zf1-4 N284R/E285F	f-A_{12}	8.6 ± 0.1	0.38 ± 0.06
Nab2p Zf1-4 D372K/K373F/K384F	f-A_{12}	7.7 ± 0.1	1.8 ± 0.3
Nab2p Zf1-4 K309A	f-A_{12}	7.2 ± 0.2	4 ± 1
Nab2p Zf1-4 K320A/R321A/K322A	f-A_{12}	6.1 ± 0.2	28 ± 9

Table 4.3. – Energetic parameters of Nab2p – RNA interactions using different RNA probes and Zf1-4 mutants.

Summing up the obtained information, the binding mode of Nab2p Zf1-4 seems to involve C-terminal helix of Zf1-2 tandem and Zf3. In addition, Zf1-2 is clearly suboptimally designed for polyadenosine recognition and a mutant restoring the residues in Zf2 in a similar way as Zf6-7 has much higher affinity. In the other hand, remodelled Zf4 (to be like Zf3) cannot cooperate with Zf3 to increase the affinity, probably because the two binding interfaces are in opposite sites of the Nab2p Zf3-4 tandem and are unable to define a continuous binding platform.

4.3. DISCUSSION

4.3.1. RNA binding mode of Nab2p Zf1-4 region

In this work the interaction between Nab2p Zf1-4 construct and RNA was characterized. It has been found that this region has modest affinity (K_D = 2.3 ± 0.2 μM) and selectivity for polyadenosine RNA, with an optimal length of 12 nucleotides.

Identification of the binding sites was performed by NMR and binding characterization of rational designed mutants. It has been established that the Zf1-4 construct has two RNA interaction interfaces: the C-terminal helix of Zf1-2 tandem and Zf3. The RNA binding to the α-helix probably is carried out through electrostatic interactions (*i.e.* unspecific), while the Zf3 binding pocket seems to share recognition mechanism with Nab2p Z5-7 (Brockmann et al., 2012) and Tis11d (Hudson et al., 2004) (*i.e.* specific).

In contrast to the Nab2p Zf5-7 region, the two tandems present in Nab2p Zf1-4 have been apparently selected to bind RNA sub-optimally. The Zf1-2 tandem has, somehow, lost RNA specificity and probably the increase of affinity of mutant in Zf2 region is due to a partial restoration of RNA specific binding modes as in Zf5-7. In contrast, in the Zf3-4 region the problem seems to be topological (different arrangement than in Zf1-2 or in Zf67), because the Zf4 hypothetically rescuing mutant is not able to increase the affinity.

Regarding the RNA direction in binding Nab2p Zf1-4, the observed changes in bound state fluorescence anisotropy values along mutants in Zf1-2 tandem and the fact that the fluorophore is attached at the 3'-end of the RNA suggest that occurs in the 3' to 5' direction. Moreover, this hypothesis is reinforced by the fact that other related protein-RNA complexes (Hudson et al., 2004; Teplova and Patel, 2008) bind RNA in the same orientation (the 5'-end of RNA to the C-terminal zinc finger).

4.3.2. RNA binding model of Nab2p

The RNA binding study of Nab2p Zf1-4 subdomain performed in this work complements the findings from previous studies on the zinc finger construct Zf5-7 and on the full-length protein. Now the affinity and optimal RNA length of Nab2p protein can be explained by the combination of Zf1-4 and Zf5-7 bindings characteristics (Figure 4.14).

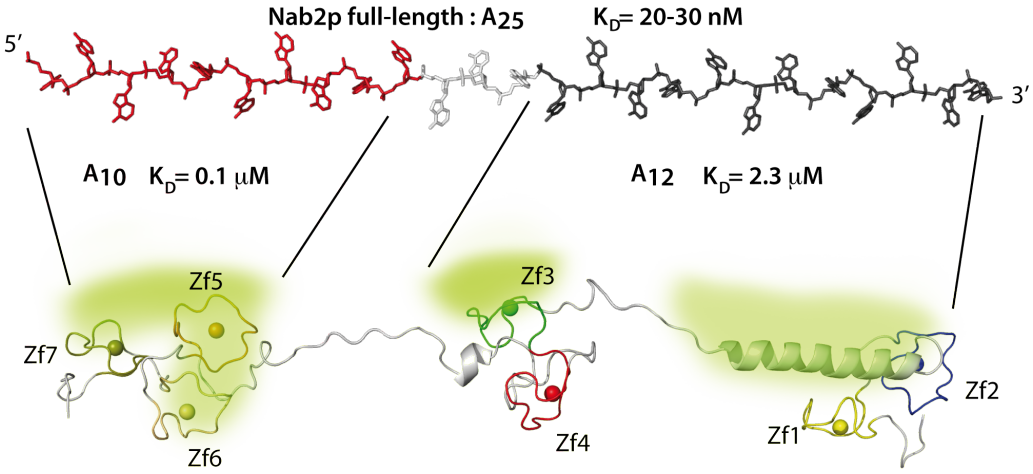


Figure 4.14. – Schematic representation of the polyadenosine RNA binding mode of Nab2p CCCH region. The molecular model has been constructed from the structures of Nab2p Zf1-2, Zf3-4 and Zf5-7; segments connecting these three structures have been built up with Pymol in an extended conformation. The RNA A_{25} represents the minimal-length RNA oligonucleotide footprinted by Nab2p and is represented in the same scale than the protein to allow the comparison between sizes.

According to this recognition model, it can be proposed a mechanism of poly (A) tail length control that shares key concepts with the mammalian model (Eckmann et al., 2011; Kuhn et al., 2009) (Figure 4.15). The nascent poly (A) tail from polymerase activity would be first bound to Nab2p Zf5-7, the higher affinity region of Nab2p (maybe in cooperation with Zf3). Afterwards, Nab2p Zf1-2 would bind weakly to the nascent tail, protecting its growth up to a length sufficient to accommodate a second Nab2p Zf5-7 unit. Once released from RNA, Nab2p Zf1-2 (or other regions of the N-terminal domains, *e.g.* the Q-rich sequence) might participate in key protein-protein interactions, perhaps in self-recognition as it happens in PABPN1, the mammalian nuclear poly (A) binding protein that regulates poly (A) length control (Eckmann et al., 2011; Kuhn and Wahle, 2004). Curiously, a systematic loss of crosspeaks in ^{1}H-^{15}N HSQC spectra corresponding to Nab2p Zf1-2 region in aged protein samples (Figure 4.15.B) and significant light scattering at 330 nm in the UV absorption spectra of the same construct were detected in previous work (Martinez-Lumbreras et al., 2013). These experimental evidences are compatible with the formation of highmolecular weight complexes (or aggregates) in which the Zf1-2 tandem seems to be involved.

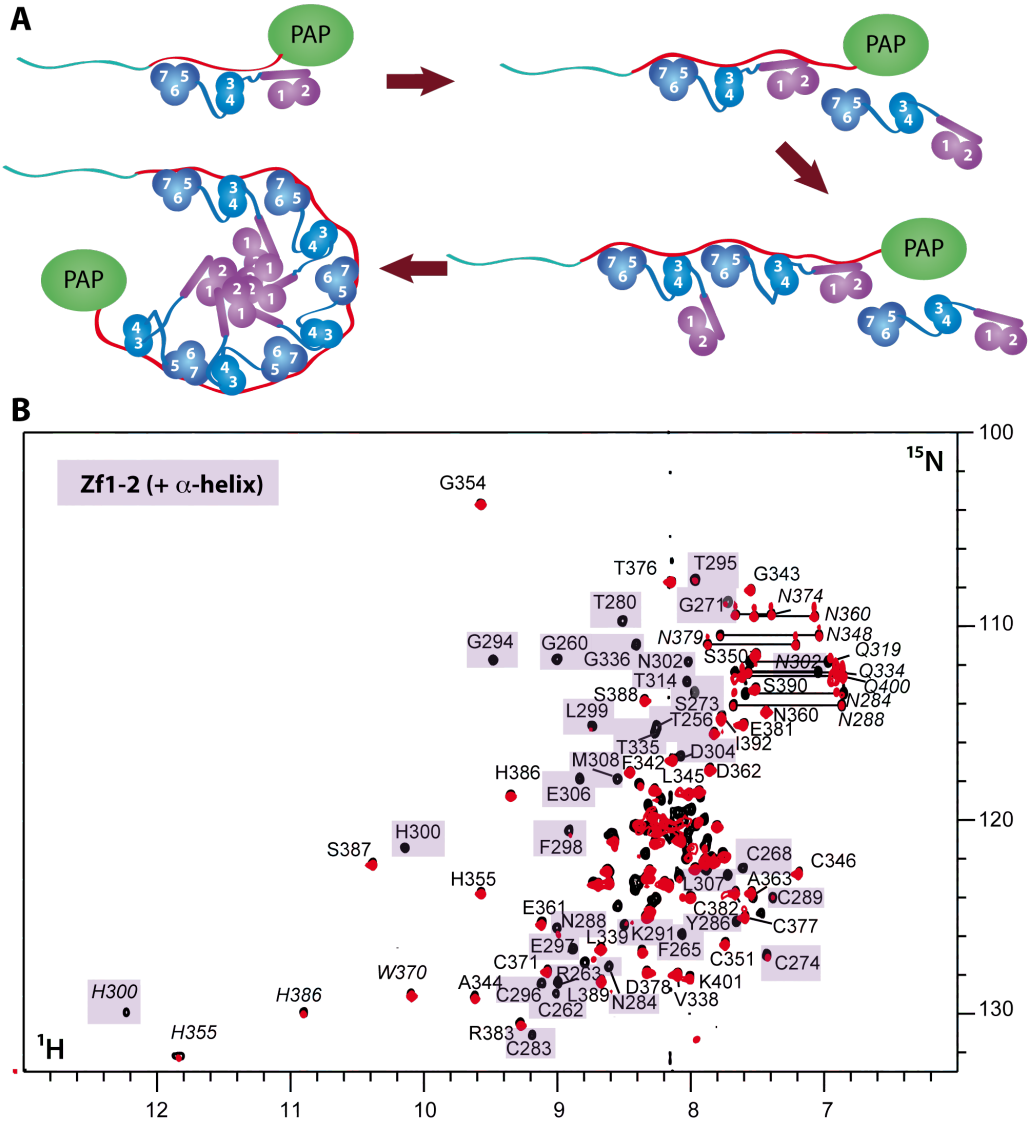


Figure 4.15. – A) Proposed mechanism of Nab2p mediated poly (A) tail length control in yeast. B) Superposition of the ^{1}H-^{15}N HSQC spectra of monomeric (black) and aggregated (red) Nab2p Zf1-4 samples. The black spectrum was obtained directly from the freshly purified sample, whereas the red one was recorded after ~ 1 month storage at -20 ºC. For clarity, only some of the assignments were included (side chains in italics) to illustrate the systematic loss of signals from the first tandem (residues 253 to 336 highlighted in purple).

4.3.3. Future perspectives

This work complements previous *in vitro* RNA binding studies of Nab2p, but it opens novel possible routes to study the poly (A) length control mediated by Nab2p. The proposed hypothesis of length control could be explored by the study of self-association of Nab2p; experimental evidences prove that Zf1-2 tandem have the ability of self-recognition (Martinez-Lumbreras et al., 2013). Thus the characterization of this self-association process either by *in vitro* and *in vivo* experiments could clarify if the control mechanism is regulated by this interaction.

In addition, some *in vivo* experiments could be performed using the designed mutant N284R/E285F to alter the RNA binding mode; this mutant is able to recognize polyadenosine RNA with much higher affinity than wild type protein, using an alternative interface in Zf1-2 tandem. To test whether this mutant controls the poly (A) tail length or not may provide new insights into this mechanism, explaining if the low RNA affinity displayed by Zf1-2 tandem has an important role in Nab2p function.

Finally, other interactions described for Nab2p could be analysed: Mex67p and Yra1p recognition regions are assigned in RGG and zinc finger domains. Using Zf1-4 construct, this interactions could be explored by NMR or by other techniques utilized along the thesis (*e.g. in vitro* cross-linking).

4.4. REFERENCES

Amrani, N., Minet, M., Le Gouar, M., Lacroute, F., and Wyers, F. (1997). Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Mol Cell Biol *17*, 3694-3701.

Anderson, J.T., Wilson, S.M., Datar, K.V., and Swanson, M.S. (1993). NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Mol Cell Biol *13*, 2730-2741.

Bergkessel, M., Wilmes, G.M., and Guthrie, C. (2009). SnapShot: Formation of mRNPs. Cell *136*, 794, 794 e791.

Brockmann, C., Soucek, S., Kuhlmann, S.I., Mills-Lujan, K., Kelly, S.M., Yang, J.C., Iglesias, N., Stutz, F., Corbett, A.H., Neuhaus, D.*, et al.* (2012). Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export. Structure *20*, 1007-1018.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835-845.

Dheur, S., Nykamp, K.R., Viphakone, N., Swanson, M.S., and Minvielle-Sebastia, L. (2005). Yeast mRNA Poly(A) tail length control can be reconstituted in vitro in the absence of Pab1p-dependent Poly(A) nuclease activity. J Biol Chem *280*, 24532-24538.

Eckmann, C.R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. Wiley Interdiscip Rev RNA *2*, 348-361.

Grant, R.P., Marshall, N.J., Yang, J.C., Fasken, M.B., Kelly, S.M., Harreman, M.T., Neuhaus, D., Corbett, A.H., and Stewart, M. (2008). Structure of the N-terminal Mlp1-binding domain of the Saccharomyces cerevisiae mRNA-binding protein, Nab2. J Mol Biol *376*, 1048-1059.

Green, D.M., Johnson, C.P., Hagan, H., and Corbett, A.H. (2003). The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. Proc Natl Acad Sci U S A *100*, 1010-1015.

Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X., and Corbett, A.H. (2002). Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. J Biol Chem *277*, 7752-7760.

Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastia, L., and Swanson, M.S. (2002). Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. EMBO J *21*, 1800-1810.

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2004). Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. Nat Struct Mol Biol *11*, 257-264.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev *24*, 1927-1938.

Kelly, S.M., Leung, S.W., Apponi, L.H., Bramley, A.M., Tran, E.J., Chekanova, J.A., Wente, S.R., and Corbett, A.H. (2010). Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3'-end formation. J Biol Chem *285*, 26022-26032.

Kelly, S.M., Pabit, S.A., Kitchen, C.M., Guo, P., Marfatia, K.A., Murphy, T.J., Corbett, A.H., and Berland, K.M. (2007). Recognition of polyadenosine RNA by zinc finger proteins. Proc Natl Acad Sci U S A *104*, 12306-12311.

Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S., and Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. J Biol Chem *284*, 22803-22814.

Kuhn, U., and Wahle, E. (2004). Structure and function of poly(A) binding proteins. Biochim Biophys Acta *1678*, 67-84.

Lemay, J.F., D'Amours, A., Lemieux, C., Lackner, D.H., St-Sauveur, V.G., Bahler, J., and Bachand, F. (2010). The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. Mol Cell *37*, 34-45.

Marfatia, K.A., Crafton, E.B., Green, D.M., and Corbett, A.H. (2003). Domain analysis of the Saccharomyces cerevisiae heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. J Biol Chem *278*, 6731-6740.

Martinez-Lumbreras, S., Santiveri, C.M., Mirassou, Y., Zorrilla, S., and Perez-Canadillas, J.M. (2013). Two Singular Types of CCCH Tandem Zinc Finger in Nab2p Contribute to Polyadenosine RNA Recognition. Structure.

Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y., and Keller, W. (1997). The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. Proc Natl Acad Sci U S A *94*, 7897-7902.

Niño, C.A., Herissant, L., Babour, A., and Dargemont, C. (2013). mRNA Nuclear Export in Yeast. Chemical reviews.

Proudfoot, N., and O'Sullivan, J. (2002). Polyadenylation: a tail of two complexes. Curr Biol *12*, R855-857.

Schmid, M., Poulsen, M.B., Olszewski, P., Pelechano, V., Saguez, C., Gupta, I., Steinmetz, L.M., Moore, C., and Jensen, T.H. (2012). Rrp6p controls mRNA poly(a) tail length and its decoration with poly(a) binding proteins. Mol Cell *47*, 267-280.

Soucek, S., Corbett, A.H., and Fasken, M.B. (2012). The long and the short of it: The role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. Biochim Biophys Acta *1819*, 546-554.

Suntharalingam, M., Alcazar-Roman, A.R., and Wente, S.R. (2004). Nuclear export of the yeast mRNA-binding protein Nab2 is linked to a direct interaction with Gfd1 and to Gle1 function. J Biol Chem *279*, 35384-35391.

Teplova, M., and Patel, D.J. (2008). Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. Nat Struct Mol Biol *15*, 1343-1351.

Tran, E.J., Zhou, Y., Corbett, A.H., and Wente, S.R. (2007). The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. Mol Cell *28*, 850-859.

Vinciguerra, P., Iglesias, N., Camblong, J., Zenklusen, D., and Stutz, F. (2005). Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. EMBO J *24*, 813-823.

Viphakone, N., Voisinet-Hakil, F., and Minvielle-Sebastia, L. (2008). Molecular dissection of mRNA poly(A) tail length control in yeast. Nucleic Acids Res *36*, 2418-2433.

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66*, 759-768.

Winstall, E., Sadowski, M., Kuhn, U., Wahle, E., and Sachs, A.B. (2000). The Saccharomyces cerevisiae RNA-binding protein Rbp29 functions in cytoplasmic mRNA metabolism. J Biol Chem *275*, 21817-21826.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiology and molecular biology reviews : MMBR *63*, 405-445.

Zheng, C., Fasken, M.B., Marshall, N.J., Brockmann, C., Rubinson, M.E., Wente, S.R., Corbett, A.H., and Stewart, M. (2010). Structural basis for the function of the Saccharomyces cerevisiae Gfd1 protein in mRNA nuclear export. J Biol Chem *285*, 20704-20715.

5. Pub1p – Tif4631p interaction

5.1. INTRODUCTION

The importance of posttranscriptional regulation of protein expression became clear with the discovery of physical and functional communication between the 5' and 3' ends of the mRNA (Gallie, 1991). Regulation at the RNA level can be achieved in several ways, including the accessibility of mRNA to translationally active (polysome) or inactive (granules) states (Balagopal and Parker, 2009). Cytoplasmic mRNA regulatory networks are largely independent of the synthesis of new mRNA transcripts and dominate gene expression control in cells with low (or null) transcription activity (*i.e.* germ cells) (Costa-Mattioli et al., 2009; Richter and Lasko, 2011) and in neural synapses than can be centimetres away from their cell nucleus (Costa-Mattioli et al., 2009; Darnell and Richter, 2012). Moreover, posttranscriptional control is a mechanism of rapid response against environmental challenges or insults to the cells (Spriggs et al., 2010). Pub1p, the protein studied in this chapter, is an essential element for the formation of the glucose deprivation RNA granules (Buchan et al., 2008; Hoyle et al., 2007; Liu-Yesucevitz et al., 2010).

5.1.1. Closed loop mRNP: translation initiation

To start the translation process, mRNAs must recruit the 43S pre-initiation complex (43S PIC), a preformed particle that contains the small ribosomal subunit (40S), some eukaryotic initiator factors (eIF1A, eIF1, eIF2, eIF3 and eIF5) and the methionylated initiator methionine transfer RNA (Met-tRNA$_i^{Met}$) (Sonenberg and Hinnebusch, 2009). The 43S PIC searches the initiation codon (AUG) and then the whole ribosome is assembled (Sachs et al., 1997). The 43S PIC recruitment is stimulated by the formation of a mRNA closed loop structure between 5' CAP binding complex and the poly (A) binding protein PABPC1 (Pab1p in yeast) attached in the 3'-end. The scaffold protein eIF4G1 (Tif4631p in yeast) bridges this interaction by simultaneously binding eIF4E (Gross et al., 2003), a protein that binds the CAP structure in 5' of mRNA, and PABPC1 (Kahvejian et al., 2001; Safaee et al., 2012) (Figure 5.1).

The closed loop structure protects mRNA from degradation and, more importantly, stimulates translation initiation by promoting 43S PIC recruitment to form 48S PIC (Figure 5.1) (Preiss and Hentze, 1999). Assemble of this complex promotes its localization near the initiation codon, which represents a crucial checkpoint in the regulation of translation initiation (Sonenberg and Hinnebusch, 2009).

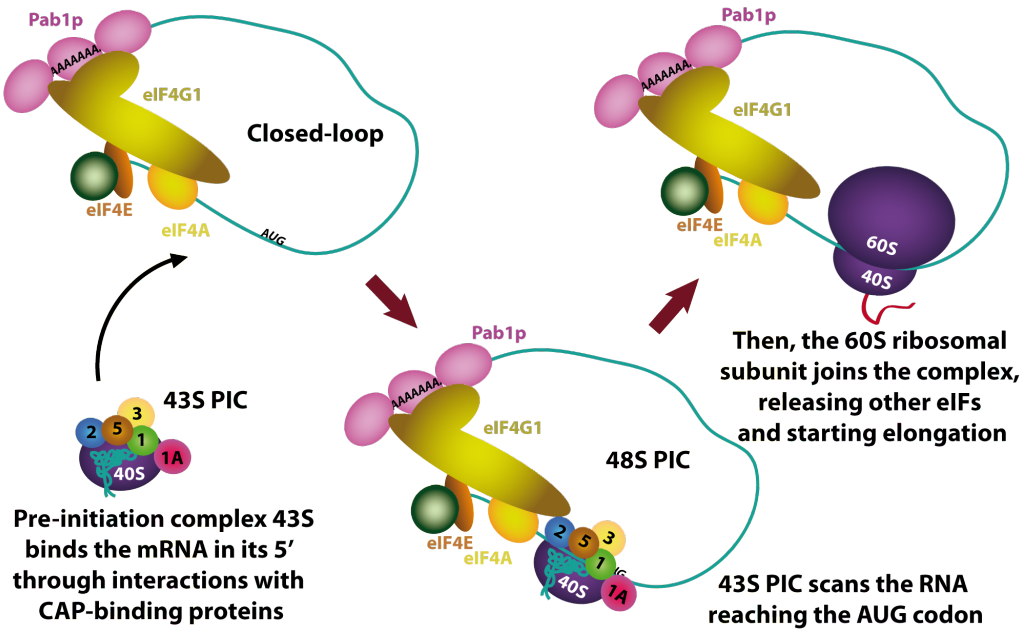


Figure 5.1. – Overview of closed loop model and of 48S PIC formation.

5.1.2. Translational arrest

Translational arrest is a universal mechanism of response to an environmental stimulus. It is regulated by cellular signalling pathways that lead to posttranslational modification of proteins involved in translation initiation (Spriggs et al., 2010). Phosphorylation of eIF2α, mainly regulated by the mTOR pathway, hinders the incorporation of the 60S subunit to the 48 PIC, effectively locking this pre-initiation complex and ultimately its accumulation (Sonenberg and Hinnebusch, 2009). Stress-induced translational arrest can be also eIF2α independent, but in all cases is characterized by the accumulation of most of the mRNAs into cytoplasmic RNA granular structures such as stress granules (SG) and processing bodies (PB), both in mammalian and in yeast cells. In general, cytoplasmic mRNPs are in a dynamic equilibrium between actively translated forms (*i.e.* polysomes) or arrested states (*i.e.* SG, PB and other RNA granules) (Figure 5.2) (Anderson and Kedersha, 2009a, b; Buchan and Parker, 2009; Moser and Fritzler, 2010; Thomas et al., 2011).

Formation of stress granules has been reported in *S. cerevisiae* upon glucose starvation (also known as EGP-bodies for eIF4E,G and Pab1p) (Brengues and Parker, 2007; Buchan et al., 2008; Hoyle et al., 2007; Swisher and Parker, 2010), robust heat shock (Grousl et al., 2009),

arsenite (Swisher and Parker, 2010), sodium azide (Buchan et al., 2011) and high ethanol levels (Kato et al., 2011). Pub1p is required for the formation of the glucose deprivation RNA granules (*i.e*. EGP-bodies) (Buchan et al., 2008; Hoyle et al., 2007; Lui et al., 2010). In addition to Pub1p, this type of RNPs contains Pab1p, Tif4631p, Tif4632p (homologs to eIF4G), Tif45p (homolog to eIF4E), Pbp1p, Ngr1p, Ygr250c, Gbp2p and Nrp1p proteins, but lack components of the 43S pre-initiation complex (including the 40S ribosomal subunit and eIF3), an important difference with the mammalian SG (Anderson and Kedersha, 2009a, b). In contrast, heat shock stress granules in yeast are compositionally more similar to mammalian SG, but their assembly is independent of eIF2α phosphorylation (Grousl et al., 2009).

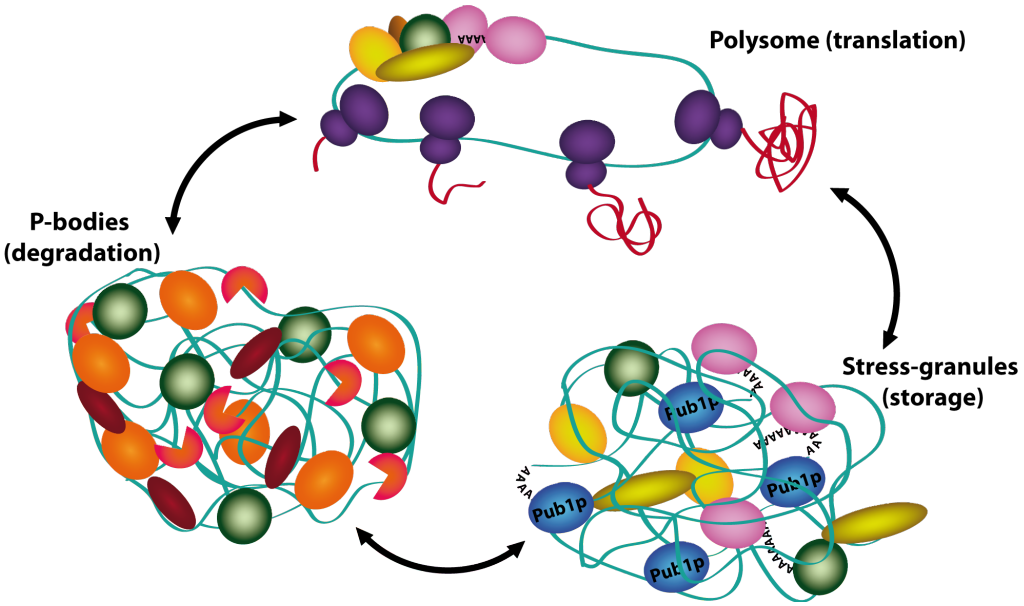


Figure 5.2. – Equilibrium between different mRNP states in cytoplasm: polysomes, stress granules and processing bodies.

Pub1p plays an important role in the formation of this cellular structures and the characterization of its interaction with other components of RNPs could clarify aspects of stress granules assembly. The interaction database shows that the stress granules components: Pab1p, Ngr1p, Nrp1p and eIF4G homologs (Tif4631p and Tif4632p) are binding partners of Pub1p. The interaction with eIF4G is particularly interesting giving the important function it plays in closed loop formation as a protein bridge between CAP and polyadenilated regions in mRNA and its potential resemblance with the recently discovered Pab1p-eIF4G interaction (Safaee et al., 2012).

5.1.3. Pub1p structure and RNA binding

S. cerevisiae Pub1p is a 453-residue protein that contains three RRM domains, two Q-rich regions (in N and C-terminal positions) and one N-rich sequence (between RRM2 and RRM3). The structures of the three RRMs have been solved: RRM1 and RRM2 have canonical RRM folds (Li et al., 2010) and RRM3 forms a novel arrangement (named TRRM, TIA1 C-terminal domain-like RRM) in which an additional N-terminal helix is packed against the canonical α_2-helix and β_2-strand (Figure 5.3) (Santiveri et al., 2011).

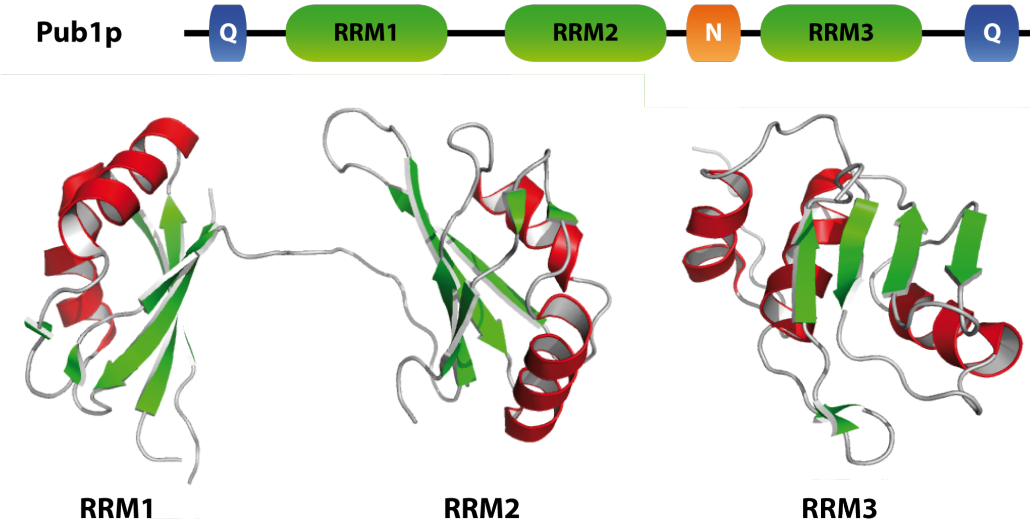


Figure 5.3. – Schematic representation of Pub1p domain organization and the structures of the RRM12 tandem (crystallographic) and the C-terminal TRRM (by NMR) (PDBs: 3MD3 and 2LA4).

Pub1p binds polyuracil RNA sequences with high affinity and selectivity principally through the RRM12 tandem (Anderson et al., 1993; Santiveri et al., 2011). The C-terminal domain also recognises RNA, showing preference for U and UA rich probes, but with less affinity than the RRM12 tandem, probably playing an auxiliary role. The binding interfaces of all domains seem to involve the canonical RNP residues in the exposed β-sheet face of each RRM (Santiveri et al., 2011).

5.1.4. Objectives

The main goal in this study is to confirm and characterize the hypothetical interaction between Pub1p and Tif4631p (eIF4G1) by biophysical/biochemical methods (mainly NMR titrations and cross-linking experiments). If confirmed, the structural aim would be to map the binding region within Pub1p and Tif4631p proteins and ultimately to characterize the structure of the complex. The final objective is to understand how the Pub1p-Tif4631p interaction might contribute to the stress granules assembly.

5.2. RESULTS

5.2.1. Tif4631p binding interface in Pub1p

The first goal in this section was to confirm the interaction between Tif4631p and Pub1p and, if so, to identify which domains are involved. Tif4631p contains binding sites for eIF4A and eIF4E in the C-terminal half and a long N-terminal region (1-402) predicted to be intrinsically unstructured that contains the Pab1p binding site (Park et al., 2011). This region of relative unknown function was the first tested for binding.

Isotopically labelled samples (^{15}N) of Pub1p RRM12 and RRM3 constructs were titrated with natural abundance Tif4631p (1-402) up to 1:2 molar ratio (Pub1p:Tif4631p) and revealed that only some signals of Pub1p RRM3 construct were perturbed, while the spectra of the RRM12 tandem remained unaltered (Figure 5.4.A). The chemical shift perturbation histograms of RRM3 (Figure 5.4.B) show small changes around a small area centred at residue Phe 366. All the signals are in fast exchange and do not seem to reach saturation values at the end of the titration suggesting that the interaction is weak.

The binding interface of Tif4631p is better described when mapping the chemical shift perturbations on the RRM3 structure (Figure 5.4.C). It involves some residues from the α_1-helix and the β_2 and β_3 strands and it is built up by the polar amino acids Gln 362 and Asn 363, whose side-chains suffer significant perturbation values, and by the hydrophobic residues Leu 360, Phe 364, Phe 366 and Leu 396.

The interaction was further confirmed by a glutaraldehyde cross-linking experiment in which Tif4631p (1-402) was mixed with Pub1p RRM12, RRM3 and with a mutant of Pub1p RRM3

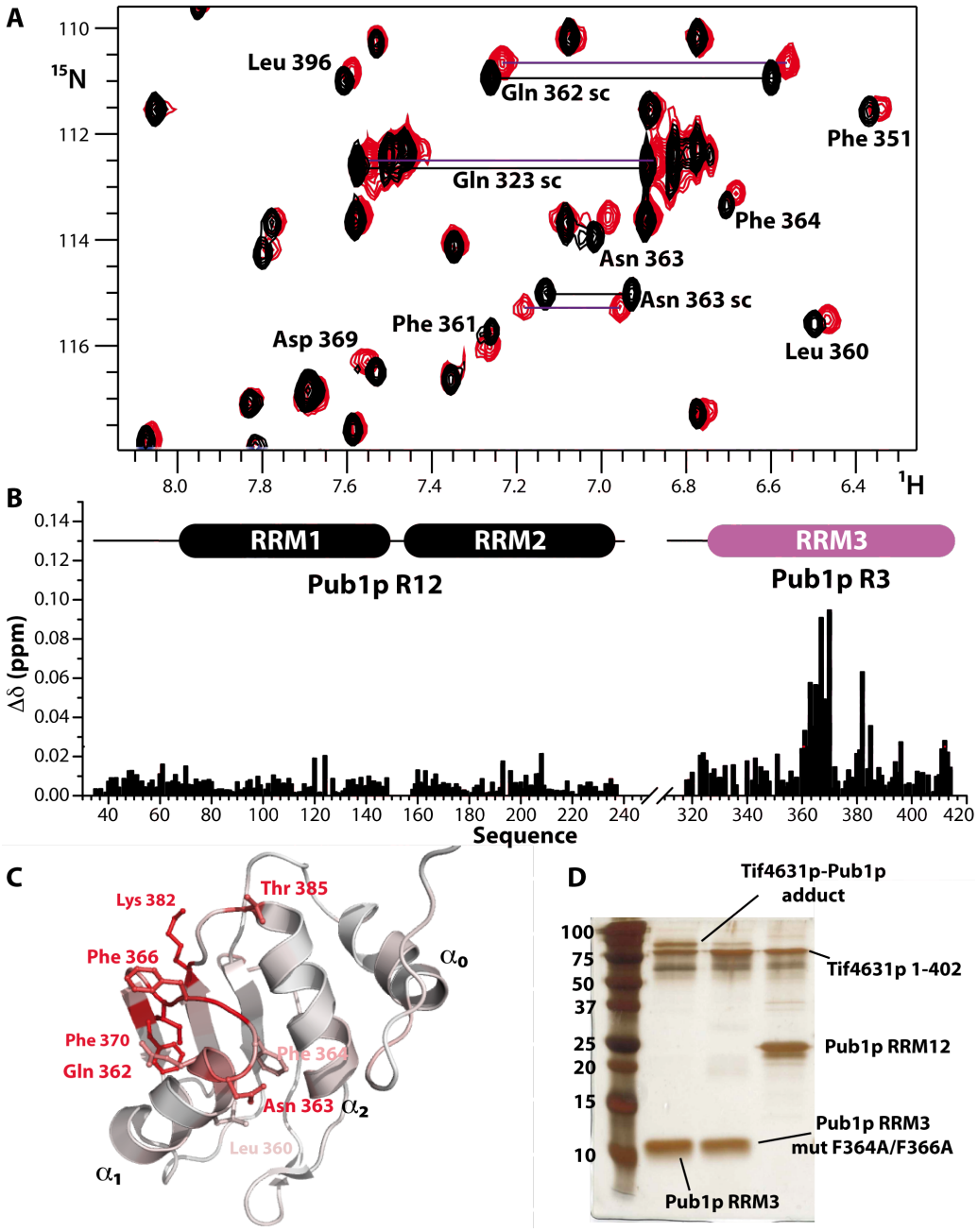


Figure 5.4. – A) Superposition of Pub1p RRM3 ^{1}H-^{15}N HSQC spectra on its free form (black) and titrated with unlabelled Tif4631p (1-402) construct (red). B) The histogram representing the averaged chemical shift perturbation values of Pub1p residues in RRM12 and RRM3 constructs (calculated following Equation 2.5). C) Chemical shift perturbation mapping on the Pub1p RRM3 structure upon titration with Tif4631p (1-402) construct. D) Silver stained SDS-PAGE of the cross-linking reaction between Tif4631p (1-402) and different proteins of Pub1p (Lane 2-RRM3 construct, Lane 3-RRM3 F364A/F366A mutant and Lane 4-RRM12 construct).

(Phe364Ala / Phe366Ala) that targets the putative binding interface (Figure 5.4.C). The reaction was performed under special conditions (see 2.11.1 section) to minimize the strong background that arises from unspecific intramolecular reactions involving the large number of reactive groups in Tif4631p (41 lysines). Protein adduct Pub1p RRM3-Tif4631p (1-402) was detected by SDS-PAGE, (Figure 5.4.D) and identified by mass spectrometry. Conversely, the cross-linking with the mutant protein render much lower adduct yield, whereas with Pub1p RRM12 no adducts were detected.

These results confirm that the interaction between Pub1p and Tif4631p occurs via the RRM3 domain and that the interface is localized between α_1-helix, β_2-strand, loop2 and loop4. The use of the helical face of RRM domains for protein-protein interactions is relatively common (Muto and Yokoyama, 2012) constituting an important peptide recognition interface compatible with the classic RNA binding interface. Moreover the results track down the interaction to a region within the first four hundred residues of Tif4631p, an important but insufficient advance to determine the chemical hallmark of this interaction. Progress in this direction required a deeper characterization of Tif4631p N-terminal region and performing further binding studies from the Tif4631p perspective as described on the next sections.

5.2.2. Characterization of Tif4631p N-terminal region

Tif4631p (eIF4G) is a scaffold protein that bridges the proteins binding the 5'-CAP and the 3' poly (A) tail, promoting mRNA circularization, ribosome recruiting and enhancing translation initiation. It is also an important component of mRNA granules in which it interacts with proteins such as Pub1p. Recognition of CAP-binding factor eIF4E and helicase eIF4A occurs at the C-terminal part of Tif4631p (Gross et al., 2003; Schutz et al., 2008); Pab1p binding sequence is located in a small box (from 197 to 224) at the N-terminal region of the protein and three different RNA interacting regions have been identified along the sequence (Berset et al., 2003). Within the N-terminal part, there are two conserved regions (Box1 and Box2) of unknown function that might constitute protein-protein interaction motifs similar to the Pab1p binding site (Box 3) (Park et al., 2011).

The Tif4631p (1-402) construct was further dissected (Figure 5.5) taking into account sequence conservation and experimental evidences about the role that the different motifs within it have for mRNP assembly (Park et al., 2011). These shorter versions progressively incorporate the motifs described above in order to evaluate their participation on Pub1p binding.

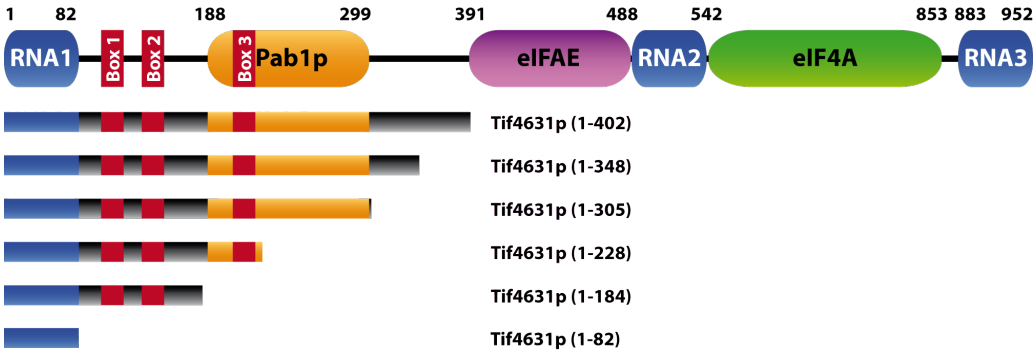


Figure 5.5. – Schematic representation of Tif4631p sequence, showing the locations of different binding regions and the constructs obtained in this work.

5.2.2.1. Long-term Tif4631p stability

The first purifications of Tif4631p constructs show multiple bands in the SDS-PAGE that were interpreted as degradation. Intrinsically unstructured proteins like Tif4631p can be quickly degraded by the presence of metallic cations in purification buffers, which either activate metaloproteases or catalyse chemical degradation (see section 2.4.4.4). Therefore purification protocols were optimized including EDTA in the buffer (to chelate traces of metallic cations) and speeding up purification to minimize the chances of degradation. Once purified, Tif4631p seems to remain stable when stored at -20 ºC or lower.

In addiction to chemical instability, some Tif4631p constructs (1-184, 1-228) showed physical instability characterized by a phase transition (liquid to gel) when the samples were stored at 4ºC during long time periods (>1 week). Formation of hydrogels has been proposed as a key process in the mechanism of nucleation of RNA granules (Kato et al., 2012) and likewise could reflect some intrinsic oligomerization tendency of Tif4631p. To examine this, Tif4631p (1-184) sequence was analysed by the zipperdatabase server [ZIPPERDB (Goldschmidt et al., 2010)] finding that peptides in Box1 and Box2 have high tendency to form fibrils (Figure 5.6.A). To test if these regions are responsible for the hydrogels formation two more Tif4631p

(1-187) variants were stored at 4ºC together with a wild type protein at similar conditions: a box1Δ (lacking residues from Pro 94 to Met 111) and a box2Δ (lacking residues from Ser 128 to His 150). About 1 week later, hydrogels appeared in wild type and in box2Δ variant, in contrast box1Δ sample remained in liquid form (Figure 5.6.B), suggesting that the Box1 region is required for Tif4631p self association. A more detailed examination of the ZIPPERDB results showed a peptide (Gly 97- Tyr 103) within Box1 with high propensity to form an amyloid-like structure (predicted model in Figure 5.6.C); the peptide is arranged in a β-strand conformation faced against another molecule of the peptide and burying one tyrosine residue in the interface region. The repetition of this pair of β-strands in the growing dimension of the fibril creates two faced parallel β-sheets (Figure 5.6.C).

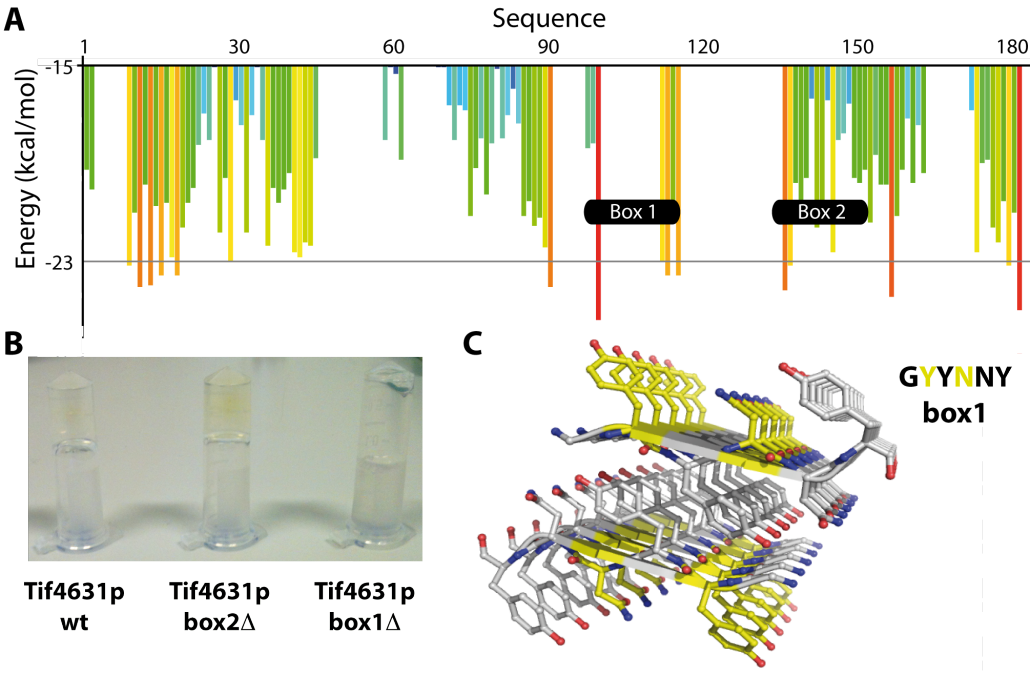


Figure 5.6. – A) Prediction of amiloyd-like sequences along Tif4631p (1-184). B) Formation of hydrogels in stored samples of purified Tif4631p (1-184) construct and mutant box2Δ, while mutant box1Δ remains liquid. C) Structural model of Box1 amyloid.

5.2.2.2. NMR analysis of Tif4631p N-terminal region

The ^{1}H-^{15}N HSQC spectra of Tif4631p (1-82, 1-184, 1-305 and 1-402) constructs were acquired under identical buffer and temperature conditions and show the characteristic of intrinsically unstructured proteins: low signal dispersion and high peak intensities. In addition, the comparison between them presented high signal overlapping (Figure 5.7).

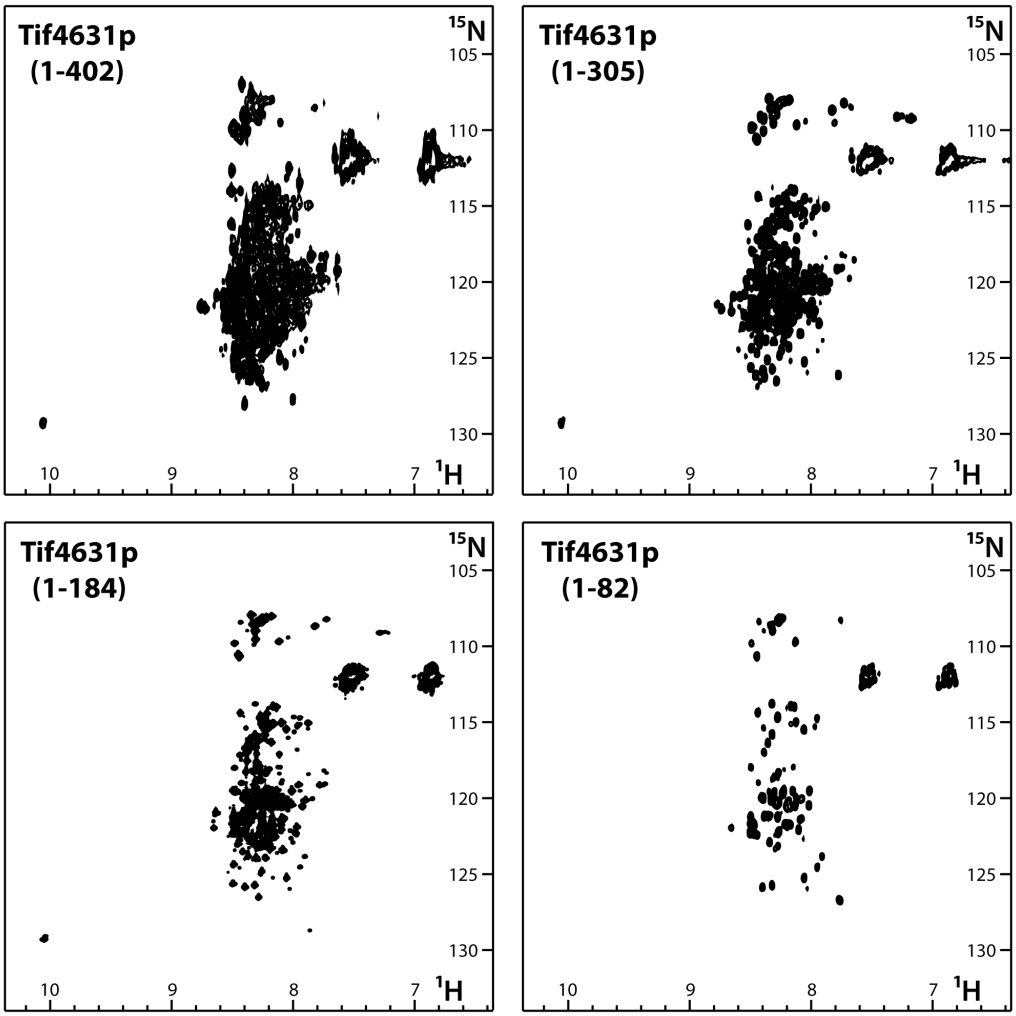


Figure 5.7. – ^{1}H-^{15}N HSQC spectra overlapping of different Tif4631p constructs, showing the poor signal dispersion, characteristic of unfolded proteins.

To perform a conformational analysis and to identify Pub1p binding region within Tif4631p, the assignment of the protein was carried out. As the 402-residue construct assignment is a hard task, a shorter version was selected (1-184), incorporating the Box1 and Box2 regions [this construct was shown to bind Pub1p RRM3 by crosslinking (see below)].

The Tif4631p (1-184) construct shows NMR spectra of better quality and allows to obtain a set of two and three-dimensional spectra: ^{1}H-^{15}N HSQC, 3D HNCO, HNCA, HN(CO)CA, HN(CA)CO, CBCANH and CBCA(CO)NH) and also two ^{13}C-detected experiments COCA and CON (Bermel et al., 2005). NMR assignment of the backbone resonances was nearly complete (93% N_H, 92% H_N, 95% CO, 96% Cα, 96% Cβ) and the CSI analysis confirmed the lack of regular secondary structure elements. The small differences to random coil values prove that protein is intrinsically unstructured, perhaps with a small α-helix population in the N and C-terminal part of the construct (Figure 5.8). Random coil values were computed considering neighbour, pH and temperature effects according with a recent set of parameters optimized for intrinsically unstructured proteins (Kjaergaard et al., 2011; Kjaergaard and Poulsen, 2011).

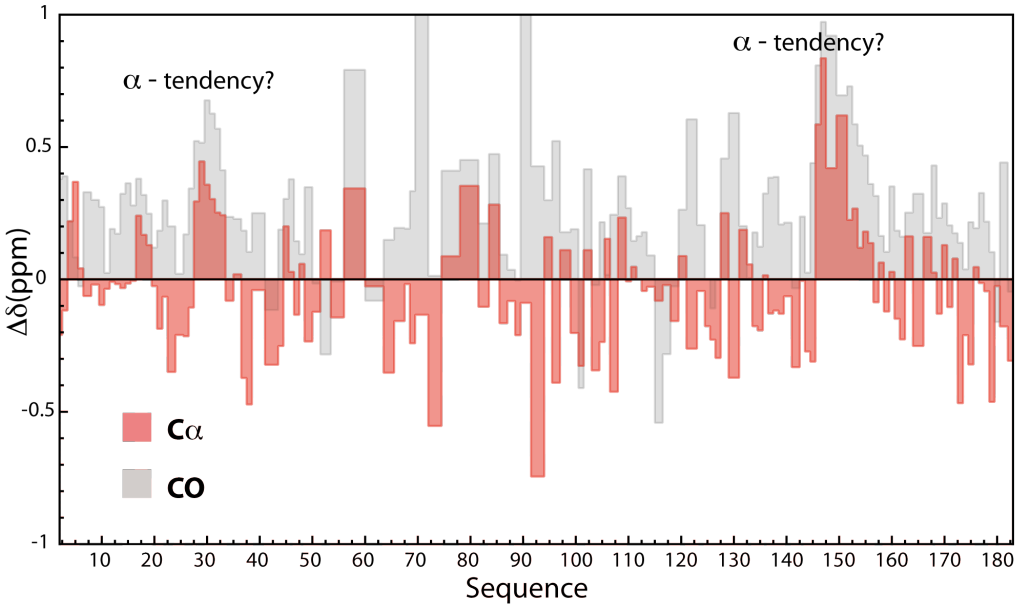


Figure 5.8. – Histograms representing the chemical shift difference between obtained Cα and CO values of Tif4631p residues and reference random coil ones.

There are further evidences suggesting that Tif4631p (1-184) construct is not totally unstructured. The presence of some amide shielded peaks in the ^{1}H-^{15}N HSQC spectra for residues Gly 42, 83 and 97, which are also broad in comparison to other Gly peaks, suggests that these regions are not totally disordered (Figure 5.9.A). The ^{13}C-detected CON spectrum is an excellent tool to analyse intrinsically unstructured proteins because it offers superior dispersion to the ^{1}H-^{15}N HSQC and optimal performance for highly dynamic NHs. Moreover, it provides the assignment for Pro backbone ^{15}N in an empty region of the spectra. All the Xxx (i-1) CO-Pro (i) N crosspeaks were assigned in the CON spectra of Tif4731p (1-184), but those belonging to Box1 appeared broad (Figure 5.9.B). This behaviour extends to other correlations in Box1 suggesting that the region has lower mobility probably due to the presence of residual structures.

And finally, another symptom of chemical instability was detected during protein assignment: deamidation of two asparagine residues, leading to two iso-aspartate residues. This chemical modification leads to new crosspeaks in the ^{1}H-^{15}N HSQC spectrum (Figure 5.9.C), which are characteristic of iso-aspartates. The two new peaks were assigned using standard 3D experiments (Appendix 7) and corresponded to Asn 41 and Asn 76 positions. Both residues are followed by Gly, a small residue that increase the reaction rate probably due to a stabilization of the pentacycle intermediate (Robinson, 2002). Deamidation is significantly slowed down at low temperatures and frozen samples remain stable for months.

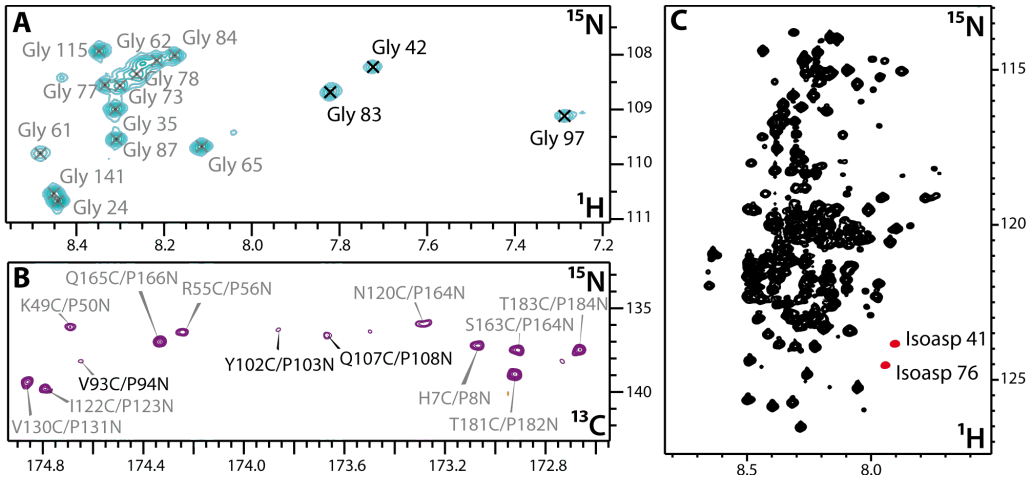


Figure 5.9. – Different spectra of Tif4631p (1-184) construct showing characteristic features. A) The different proton chemical shift of some Gly NH correlations in ^{1}H-^{15}N HSQC spectra. B) Xxx-Pro correlation region in CON spectra. C) Isoaspartate NH correlation in ^{1}H-^{15}N HSQC spectra (highlighted in red).

5.2.3. Analysis of Pub1p binding in Tif4631p

The experiments described above (5.2.1 section) revealed that Tif4631p (1-402) construct is able to bind Pub1p RRM3. In order to better delimitate the region that interacts with Pub1p, additional cross-linking and NMR titrations experiments were carried out.

5.2.3.1. Heterobifunctional cross-linking experiments

The different Tif4631p constructs were cross-linked with Pub1p RRM3 but in this case, using a more specific heterobifuntional reactant (sulfo-NHS-LC-diazirine) instead of glutaraldehyde. The protocol, described in 2.12.2 section, consisted of an activation step where the cross-linker is mixed with Pub1p RRM3 protein and then, after eliminating the reactant excess, the cross-linking reaction of activated Pub1p with the different Tif4631p constructs under UV light. The method offers better performance than previous assays, with less background.

The reactions showed that all constructs were able to bind Pub1p RRM3 and to form the corresponding adduct products (Figure 5.10). In the case of Tif4631p (1-82) construct (corresponding to the first RNA binding region of Tif4631p) only a faint adduct band was observed, thus it was initially discarded for further analyses. The remaining constructs contain Box1 and Box2 sequences whereas the region containing the Pab1p binding site (Box3) seems to be dispensable for Pub1p binding.

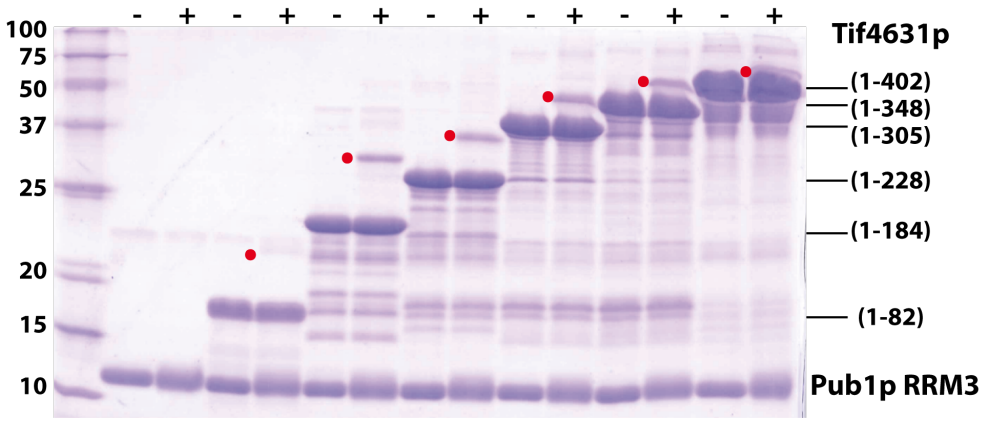


Figure 5.10. – SDS-PAGE of different cross-linking reactions. Pub1p RRM3 (lower band of each lane) in activated form (+) and without cross-linker (-) is mixed with different Tif4631p constructs. Adduct bands are marked with a red dot.

5.2.3.2. NMR titrations using different Tif4631p constructs

On the next step, the interaction was analysed by looking at changes on the ^{1}H-^{15}N HSQC spectrum of Tif4631p (1-184) caused upon titration with unlabelled Pub1p RRM3. Only a few peaks suffer changes that are small but reproducible in a longer construct [Tif4631p (1-305)], indicating that the interaction seem to be weak (Figure 5.11.A). The perturbed signals are distributed along three different segments in the Tif4631p sequence (Figure 5.11.B): two regions around positions 40 and 70 and a third one coinciding with Box1. In contrast, resonances from the second box remain unperturbed. The three perturbed regions show high degree of homology and it is possible to propose a binding motif (YNN---Y) (Figure 5.11.C).

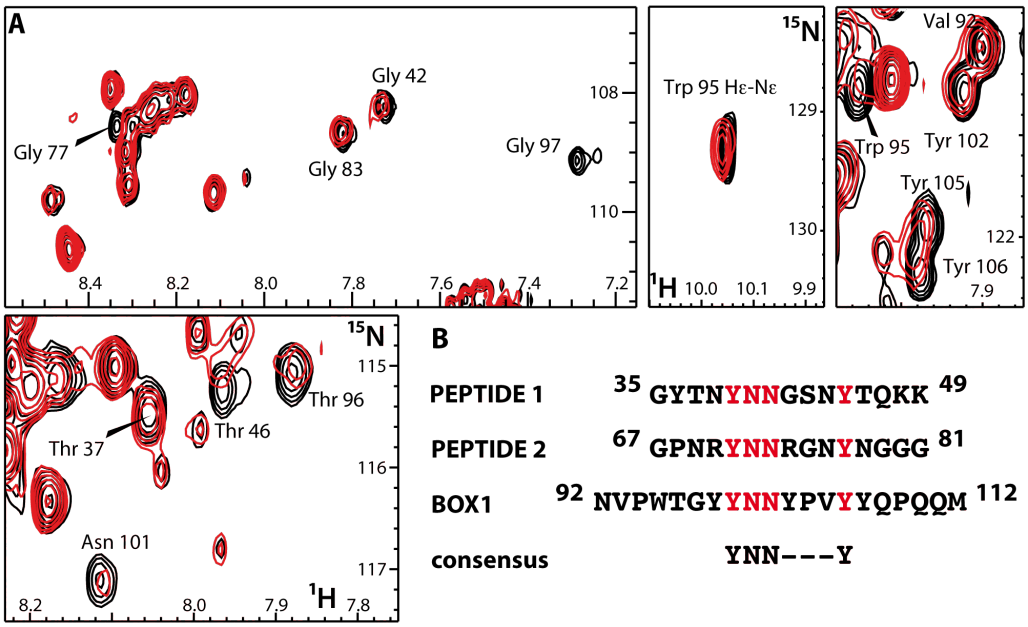


Figure 5.11. – A) Details of ^{1}H-^{15}N HSQC spectra comparison between free Tif4631p (1-184) construct and bound form to Pub1p RRM3. B) Sequence alignment of the three regions involved in Pub1p recognition.

Surprisingly the motif coincides with that postulated as a putative amiloydogenic sequence in the Box1 region and in addition the two asparagine deamidations occurs within Peptide 1 and Peptide 2. The result contrast with the chemical cross-linking data that discarded the region 1-82 as Pub1p RRM3 binder due to the weakness of the hypothetical adduct band. The reason behind this discrepancy might be in the special behaviour of the Tif4631p (1-82) construct that is poorly stained with coomassie (even negatively stained with silver). Alternatively, it could happen that changes in the NMR spectra of Tif4631p are indirect (*e.g.* conformational

rearrangement, loosing long-range contacts.) and not due to direct Pub1p binding. To clarify the binding ability to Pub1p RRM3 of several peptides corresponding to the potential binding sites were assayed.

5.2.4. Peptide titrations in Pub1p RRM3

NMR titrations were performed using labelled Pub1p RRM3 protein and three unlabelled synthetic peptides covering the putative binding regions of Tif4631p (Figure 5.11.B). The results of these titrations (Figure 5.12.A) showed that Peptide 1 and Box1 bind Pub1p using the same interface as Tif4631p (1-402), whereas Peptide 2 does not present significant changes on the ^{1}H-^{15}N HSQC spectrum and therefore does not bind. Peptide 1 produces larger shift perturbations than Box1 but this later presents a more extended interface including some residues from the α_0-helix (Figure 5.12.B).

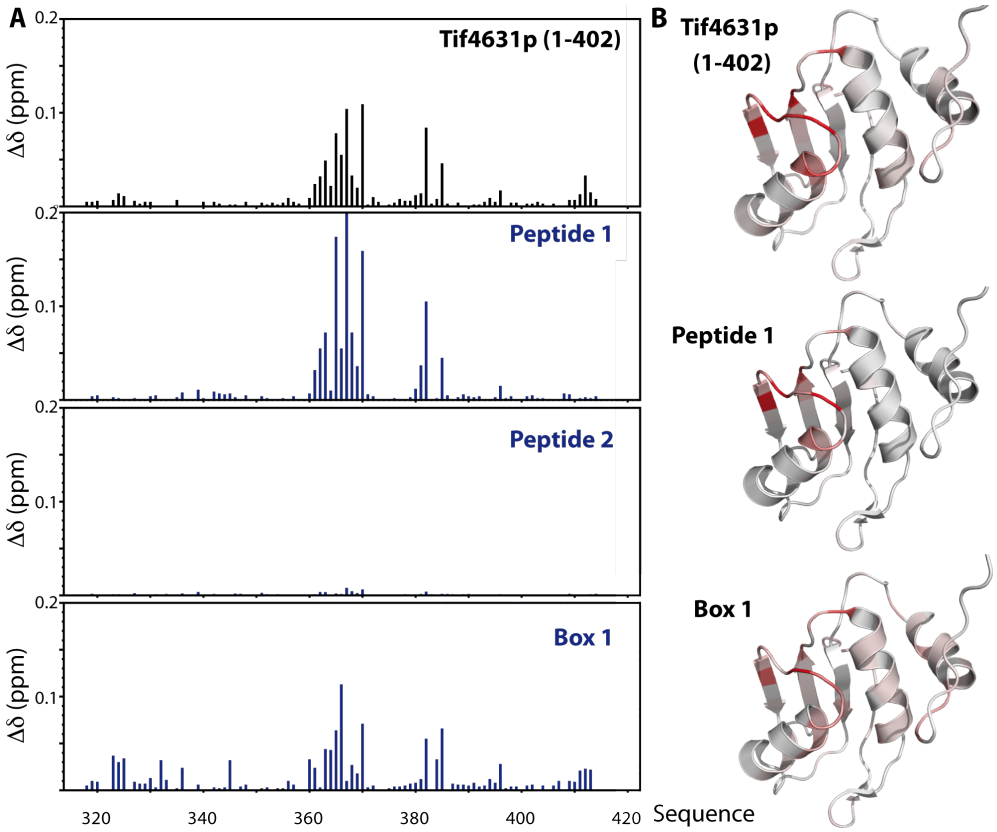


Figure 5.12. – A) Histograms of Pub1p RRM3 chemical shift perturbation upon titration of different Tif4631p peptides compared with the Tif4631p (1-402) construct one. B) Mapping on Pub1p RRM structure of the perturbation residues upon Box1 titration.

The sequence variations among the three peptides (Figure 5.11.B) were examined to explain their different binding behaviour. The most remarkable difference between Peptide 2 (non binder) and the other two is the presence of two arginine residues flanking the YNN motif; that are two tyrosines in Box1, and an asparagine and a glycine in Peptide 1. The positive charge in Peptide 2 could be responsible for the lack of binding with Pub1p, although no positive (neither negative) charged residues appear along its RRM binding interface. By contrast, the extension of the interacting platform with Box1 peptide suggests that the common binding interface recognises the YNN motif (common for both peptides) and the α_0-helix region should recognize other amino acids of Box1. A more detailed examination of the chemical shift changes highlighted some signals with distinct behaviours. The side chain amides of Gln 388 and 362 are among the most affected by binding, however they move differently in Peptide 1 and Box1, probably as a consequence of their different chemical environments (Figure 5.13).

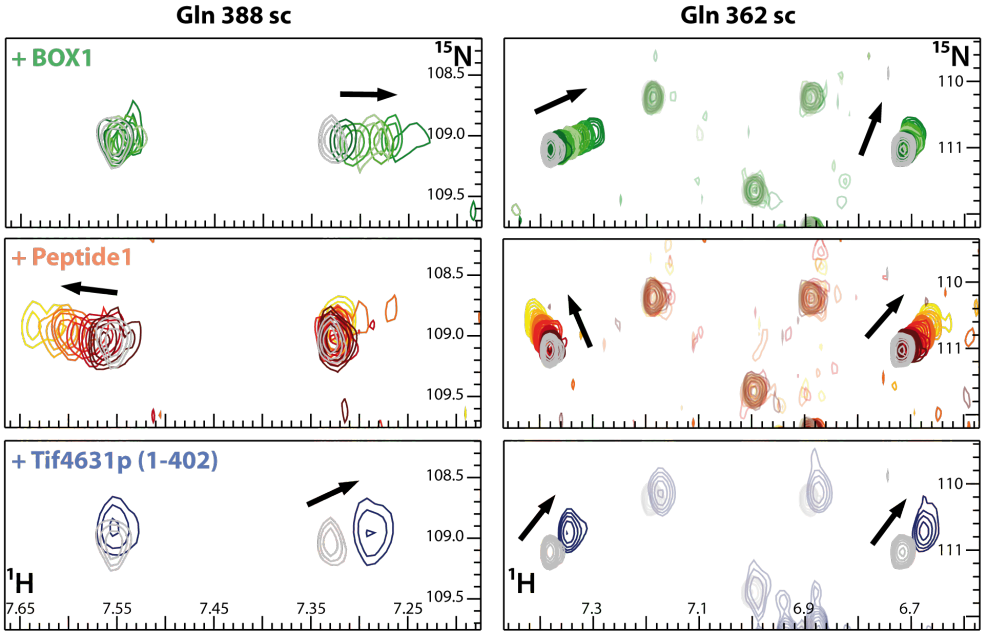


Figure 5.13. – ^{1}H-^{15}N HSQC details of Gln 388 and 362 side chain signals upon different titrations with Box1 sequence (up in green), Peptide 1 (middle in orange) and with Tif4631p (1-402) protein (down in blue). In all of them free Pub1p RRM3 spectrum is highlighted in grey.

5.2.5. Tif4631p induces Pub1p aggregation

The identification of two Pub1p binding peptides opens the question about which of the two dominate in the context of Tif4631p. This is difficult to address by looking at the Tif4631p ^{1}H-^{15}N HSQC, but should be easier to address looking at the Pub1p RRM3 one (Figure 5.13). The side chain signals of Gln 388 and Gln 362 (and others) change in the Tif4631p (1-402) with a direction between Peptide 1 and Box1 titrations suggesting that the two sites are populated, maybe been preferentially bound to Box1. The same pattern of changes is shown for Tif4631p (1-348) and (1-305) confirming the absence of Pub1p binding sites in the segment 300-400.

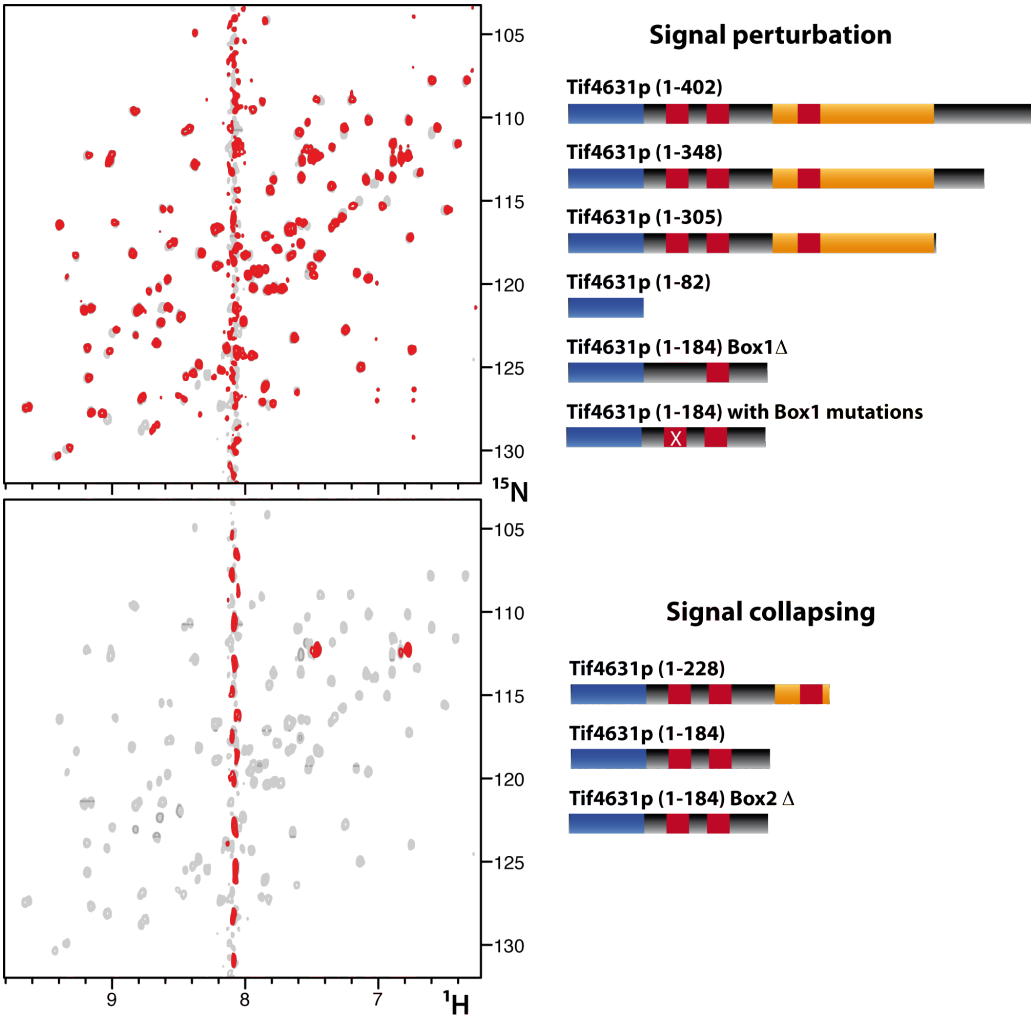


Figure 5.14. – Different ^{1}H-^{15}N HSQC spectra of labelled Pub1p RRM3 construct titrated with several Tif4631p constructs and mutants (red), overlapped with the free spectrum (grey).

Surprisingly the titrations with Tif4631p (1-228) and (1-184) caused a complete loss of the Pub1p RRM3 signals in ^{1}H-^{15}N HSQC spectra (Figure 5.14), suggesting protein aggregation (a similar behaviour as Gbp2p RRM12 upon TG-43 binding; 3.2.5.1.4 section). The weak binding situation is recovered when titrating with Tif4631p (1-82) and this time the changes resembles more those caused by Peptide 1 (but not completely).

The aggregation behaviour is interesting and was further investigated by looking to titration experiments with some Tif4631p (1-184) mutants. Those affecting the composition of Box1 (box1Δ, W95A and Y105A/Y106A) prevent the formation of Pub1p aggregates whereas Tif4631p 1-184 box2Δ causes aggregation (Figure 5.14). The results resemble those of the hydrogels (section 5.2.2.1) suggesting that both effects could have similar molecular origins.

These experiments prove that the Box1 region in Tif4631p is involved in some kind of aggregation process in which Pub1p might participate by being captured in the aggregated form. But the absence of this signal collapsing in the larger constructs (1-305, 1-348 and 1-402) suggests that this process is more complicated and interactions between different Tif4631p regions might prevent aggregation. This and other aspects will be discussed in the next section.

5.3. DISCUSSION

5.3.1. Tif4631p is a protein hub in translation regulatory networks

Intrinsically unstructured proteins (IUPs) represent about 30% of the eukaryote genomes (Ward et al., 2004). NMR is an excellent technique to study them as provides information at residue level. This type of proteins usually plays coordination roles in interaction regulatory networks, in fact, most of the hub proteins have large intrinsically unstructured regions (Gsponer and Babu, 2009). The IUPs have unique molecular recognition features: they bind through short sequence motifs (7-22 residues) also known as MoRFs (Molecular recognition features) and their binding follows fold-upon-binding mechanism characterized by a high entropy costs (large loss of degrees of freedom) and high enthalpy gains (large interfaces). These two terms contribute in different directions to the Gibbs energy; therefore the complexes formed by IUPs are characterized by low affinities but high specificities. The high degree of adaptability of the unfolded protein causes promiscuous recognition: one protein

can bind several partners, often using overlapping motifs. Finally, most posttranslational modification motifs appear within unstructured regions.

Tif4631p is a classic IUP involved in mRNA regulatory networks. More than 50% of the protein is intrinsically unstructured on its native state. Previous NMR studies on the segment 348 to 513 shows that is unstructured and acquired a α-helical fold upon binding to eIF4E (Ptushkina et al., 1998), the CAP-binding protein. The current study reveals that the first 402 residues are also unstructured on the native stage and provides a nearly complete NMR assignment for the 1-184 segment. In addition, the construct Tif4631 (1-402) contains several conserved motifs that can be considered as MoRFs:

- Box 3 (residues 197 to 224) defines the binding site of Pab1p. The recent structure of the human homolog PABPC1 domains RRM1 and RRM2 reveals the molecular details of the interaction with the second RRM and the peptide segment of eIF4G1. The later folds into α/β structure and interacts with the β_4-strand of the RRM.
- Box 2 (residues 128 to 150) is conserved in fungal eIF4Gs but so far no binding partner has been found.
- Box 1 (residues 94 to 111) interacts with Pub1p RRM3 and is required for hydrogel formation (probably forming amyloid like structures).
- RNA binding region 1 (1-82). It is one of the three regions in Tif4631p that has been proven to bind RNA experimentally. Two MoRFs, corresponding to segments 35-49 (Peptide 1) and 67-81 (Peptide 2), have been identified in this work.

Peptide 1, Peptide 2 and Box1 have a characteristic motif (YNN---Y, see 5.2.3.2 section) and NMR binding data suggest that Peptide 2 does not bind Pub1p. Therefore the changes observed in the Tif4631p spectra for the region around Gly 77 should not be attributable to direct contacts to Pub1p RRM3, but rather to some rearrangements of intermolecular contacts involving the Peptide 2 region.

On the other hand, Box1 has been proved to be key for the oligomerization/aggregation process. Pub1p might speed up the formation of these aggregates, however the presence of an undetermined element between positions 228 and 305 in Tif4631p abrogates this oligomerization mechanism, perhaps by intramolecular interactions with Box1. Curiously this region has a NNN sequence, which resembles to the YNN motif.

Aging of Tif4631p causes a macroscopic effect (hydrogel formation) and a chemical modification, the deamidation of Asn 41 and Asn 71, that are located in Peptide 1 and Peptide 2 respectively. Since deamination has been linked to amyloid formation (Dunkelberger et al., 2012), it is interesting to speculate about the possibility that both aging effects in Tif4631p are linked as well. Deamidation might accelerate aggregation by, for example, destabilizing Box1-Peptide 2 contacts (see above).

These hypothetical contacts can be integrated in a tentative model that could help understanding the role of Tif4631p as hub protein (Figure 5.15). The oligomerization mediated by Box 1 would be at the hearth of this hub. Several intramolecular contacts would prevent aggregation by competing with the nucleation process. These safeguards are simultaneously binding sites for proteins as Pub1p and (perhaps) Pab1p, which could interfere with intramolecular contacts and promote aggregation. Working on the same direction aging effects would also favour Box1-mediated aggregation.

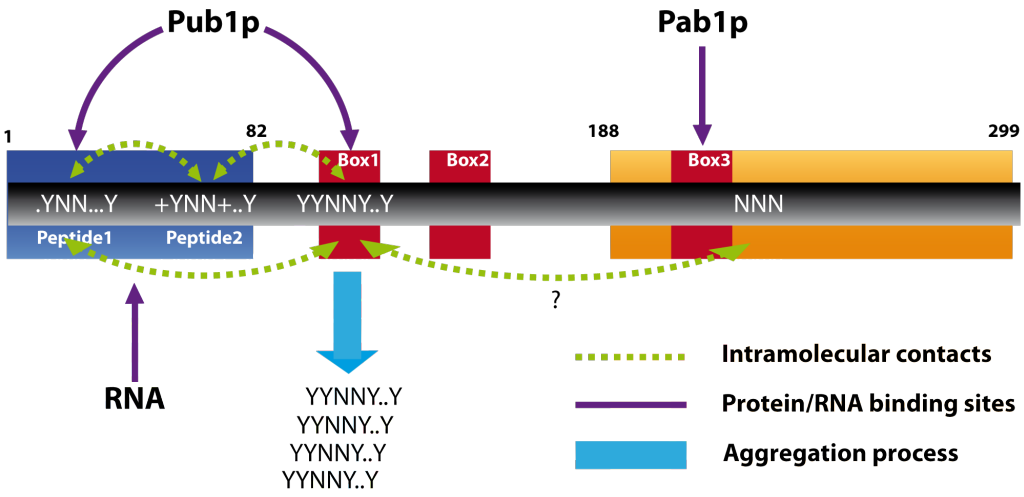


Figure 5.15. – Model of the Tif4631p different equilibrium: aggregation process directed by Box 1, the different intramolecular associations that protect the aggregation process and the role that Pub1p and Pab1p could play in these processes.

This work identified two MoRFs in Tif4631p that specifically bind to Pub1p RRM3. This interaction resembles that in the human eIF4G-PAB1 complex, probably similar in yeast (Pab1p RRM2/Tif4631p Box3). In both cases the protein-protein interfaces are opposite to the RNA binding ones, but the precise locations suggest that recognition modes are different. Tif4631p binding interface in Pub1p resides in the groove between the canonical α_1-helix and

the β_2-strand of the C-terminal RRM also involving residues from loops 2 and 4 (Figure 5.4), although it could be expanded incorporating some amino acids from the α_0-helix, as suggested by Box1 peptide titrations (Figure 5.12). In contrast PAB1 RRM2 hosted the eIF4G peptide between the two canonical α-helixes. The unstructured peptide folds displaying a β-hairpin structure upon binding with PAB1, as occurs in many MoRFs. By analogy, the same reorganization could happen in Tif4631p-Pub1p binding. But, confirmation of this last proposal would require more information (Pub1p RRM3-Tif4631p complex).

5.3.2. Implications for stress granules assembly

Pub1p is loaded in the nuclear mRNP and then exported to the cytoplasm, where is part of some translationally inactive mRNPs (Anderson et al., 1993). The way in which this protein is removed from the active polysomes is not known; it could involve some releasing step in the translational initiation (Sonenberg and Hinnebusch, 2009) or perhaps during the pioneer round of translation (Maquat et al., 2010). Different environmental stressors could alter the correct translation initiation process leading into accumulation of pre-initiation complexes and forming the different cytoplasmic granules. In humans, the Pub1p homolog (TIA-1) participates actively in this aggregation involving its C-terminal prion-like region (Gilks et al., 2004; Kedersha et al., 1999). By homology, the glutamine rich region in the N and C-terminal parts of Pub1p and/or the asparagine rich sequence between domains RRM2 and RRM3 could cause the accumulation of those PICs that still contain Pub1p, triggering the formation of RNA granules.

In this work, however, it has been revealed that Tif4631p contains a sequence that aggregates *in vitro* and could participate in pre-initiation complexes accumulation. Moreover, Pub1p binding to this (or near this) sequence triggers a sudden collapsing of Pub1p RRM3 into large aggregates. This *in vitro* interaction could have important implications in the *in vivo* formation of cytoplasmic granules under environmental stress, where Pub1p bound to Tif4631p causes its accumulation. In addition, the Q and N rich regions could help to the formation of these aggregates similar as in TIA-1 protein.

Other biomolecules could affect the aggregation equilibrium of Tif4631p: Peptide 1 and Peptide 2 are located in the RNA binding region, and the NNN sequence is near Pab1p binding site. The binding of RNA or/and Pab1p could also induce Box1 exposure, triggering the

aggregation. But these hypotheses are quite speculative and need further experiments to be confirmed. First, it will be needed to prove that Box1 sequence of Tif4631p is required for *in vivo* granules formation and then to explore the relation of different partners in binding and aggregation.

5.3.3. Future perspectives

Some NMR experiments using paramagnetic probes placed along the protein sequence could be done to test the existence of intramolecular contacts in Tif4631p. It would be clarifying to know if the three YNN---Y motifs (Peptide 1, Peptide 2 and Box 1) contact between them and also to discover if other sequences participate in this interaction (*e.g.* the NNN motif in Pab1p binding site).

The hypothetical relationship between Tif4631p intramolecular contacts and aggregation equilibrium should be explored more systematically, including the effect that Pub1p (and Pab1p) may play in this. For this, an exhaustive study of the interaction network of the Tif463p N-terminal region could be done using the approach followed in this work (NMR and cross-linking). Different mixtures of Pub1p, Pab1p, Tif4631p and RNA (and the variety of mutants designed in this work) will reveal if simultaneous binding can occur, and also it would explain the different circumstances that trigger the aggregation process.

Finally, all the knowledge generated could be used as guidelines to design and perform different *in vivo* experiments with *S. cerevisiae* in order to fully understand the structure-function relationships, for example, it could be analysed the formation of stress granules in a variety of deletion mutants (Tif4631p Box1Δ, Pub1p F364A/F366A,…).

5.4. REFERENCES

Anderson, J.T., Paddy, M.R., and Swanson, M.S. (1993). PUB1 is a major nuclear and cytoplasmic polyadenylated RNA-binding protein in Saccharomyces cerevisiae. Molecular and cellular biology *13*, 6102-6113.

Anderson, P., and Kedersha, N. (2009a). RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat Rev Mol Cell Biol *10*, 430-436.

Anderson, P., and Kedersha, N. (2009b). Stress granules. Curr Biol *19*, R397-398.

Balagopal, V., and Parker, R. (2009). Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. Current opinion in cell biology *21*, 403-408.

Bermel, W., Bertini, I., Felli, I.C., Piccioli, M., and Pierattelli, R. (2005). 13C-detected protonless NMR spectroscopy of proteins in solution. Progress in Nuclear Magnetic Resonance Spectroscopy *48*, 21.

Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. RNA (New York, NY *9*, 871-880.

Brengues, M., and Parker, R. (2007). Accumulation of polyadenylated mRNA, Pab1p, eIF4E, and eIF4G with P-bodies in Saccharomyces cerevisiae. Mol Biol Cell *18*, 2592-2602.

Buchan, J.R., Muhlrad, D., and Parker, R. (2008). P bodies promote stress granule assembly in Saccharomyces cerevisiae. J Cell Biol *183*, 441-455.

Buchan, J.R., and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. Molecular cell *36*, 932-941.

Buchan, J.R., Yoon, J.H., and Parker, R. (2011). Stress-specific composition, assembly and kinetics of stress granules in Saccharomyces cerevisiae. J Cell Sci *124*, 228-239.

Costa-Mattioli, M., Sonenberg, N., and Richter, J.D. (2009). Translational regulatory mechanisms in synaptic plasticity and memory storage. Progress in molecular biology and translational science *90*, 293-311.

Darnell, J.C., and Richter, J.D. (2012). Cytoplasmic RNA-binding proteins and the control of complex brain function. Cold Spring Harbor perspectives in biology *4*, a012344.

Dunkelberger, E.B., Buchanan, L.E., Marek, P., Cao, P., Raleigh, D.P., and Zanni, M.T. (2012). Deamidation accelerates amyloid formation and alters amylin fiber structure. Journal of the American Chemical Society *134*, 12658-12667.

Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes & development *5*, 2108-2116.

Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoecklin, G., Dember, L.M., and Anderson, P. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. Mol Biol Cell *15*, 5383-5398.

Goldschmidt, L., Teng, P.K., Riek, R., and Eisenberg, D. (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proceedings of the National Academy of Sciences of the United States of America *107*, 3487-3492.

Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., McCarthy, J.E., and Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell *115*, 739-750.

Grousl, T., Ivanov, P., Frydlova, I., Vasicova, P., Janda, F., Vojtova, J., Malinska, K., Malcova, I., Novakova, L., Janoskova, D., *et al.* (2009). Robust heat shock induces eIF2alpha-phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, Saccharomyces cerevisiae. J Cell Sci *122*, 2078-2088.

Gsponer, J., and Babu, M.M. (2009). The rules of disorder or why disorder rules. Progress in biophysics and molecular biology *99*, 94-103.

Hoyle, N.P., Castelli, L.M., Campbell, S.G., Holmes, L.E., and Ashe, M.P. (2007). Stress-dependent relocalization of translationally primed mRNPs to cytoplasmic granules that are kinetically and spatially distinct from P-bodies. J Cell Biol *179*, 65-74.

Kahvejian, A., Roy, G., and Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. Cold Spring Harbor symposia on quantitative biology *66*, 293-300.

Kato, K., Yamamoto, Y., and Izawa, S. (2011). Severe ethanol stress induces assembly of stress granules in Saccharomyces cerevisiae. Yeast *28*, 339-347.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J.*, et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Kedersha, N.L., Gupta, M., Li, W., Miller, I., and Anderson, P. (1999). RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. J Cell Biol *147*, 1431-1442.

Kjaergaard, M., Brander, S., and Poulsen, F.M. (2011). Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. Journal of biomolecular NMR *49*, 139-149.

Kjaergaard, M., and Poulsen, F.M. (2011). Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. Journal of biomolecular NMR *50*, 157-165.

Li, H., Shi, H., Wang, H., Zhu, Z., Li, X., Gao, Y., Cui, Y., Niu, L., and Teng, M. (2010). Crystal structure of the two N-terminal RRM domains of Pub1 and the poly(U)-binding properties of Pub1. J Struct Biol *171*, 291-297.

Liu-Yesucevitz, L., Bilgutay, A., Zhang, Y.J., Vanderwyde, T., Citro, A., Mehta, T., Zaarur, N., McKee, A., Bowser, R., Sherman, M.*, et al.* (2010). Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. PLoS One *5*, e13250.

Lui, J., Campbell, S.G., and Ashe, M.P. (2010). Inhibition of translation initiation following glucose depletion in yeast facilitates a rationalization of mRNA content. Biochem Soc Trans *38*, 1131-1136.

Maquat, L.E., Tarn, W.Y., and Isken, O. (2010). The pioneer round of translation: features and functions. Cell *142*, 368-374.

Moser, J.J., and Fritzler, M.J. (2010). Cytoplasmic ribonucleoprotein (RNP) bodies and their relationship to GW/P bodies. Int J Biochem Cell Biol *42*, 828-843.

Muto, Y., and Yokoyama, S. (2012). Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems. Wiley interdisciplinary reviews RNA *3*, 229-246.

Park, E.H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1*PABP mRNPs in vivo. The EMBO journal *30*, 302-316.

Preiss, T., and Hentze, M.W. (1999). From factors to mechanisms: translation and translational control in eukaryotes. Curr Opin Genet Dev *9*, 515-521.

Ptushkina, M., von der Haar, T., Vasilescu, S., Frank, R., Birkenhager, R., and McCarthy, J.E. (1998). Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20. The EMBO journal *17*, 4798-4808.

Richter, J.D., and Lasko, P. (2011). Translational control in oocyte development. Cold Spring Harbor perspectives in biology *3*, a002758.

Robinson, N.E. (2002). Protein deamidation. Proc Natl Acad Sci U S A *99*, 5283-5288.

Sachs, A.B., Sarnow, P., and Hentze, M.W. (1997). Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell *89*, 831-838.

Safaee, N., Kozlov, G., Noronha, A.M., Xie, J., Wilds, C.J., and Gehring, K. (2012). Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. Molecular cell *48*, 375-386.

Santiveri, C.M., Mirassou, Y., Rico-Lastres, P., Martinez-Lumbreras, S., and Perez-Canadillas, J.M. (2011). Pub1p C-terminal RRM domain interacts with Tif4631p through a conserved region neighbouring the Pab1p binding site. PLoS One *6*, e24481.

Schutz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America *105*, 9564-9569.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational regulation of gene expression during conditions of cell stress. Molecular cell *40*, 228-237.

Swisher, K.D., and Parker, R. (2010). Localization to, and effects of Pbp1, Pbp4, Lsm12, Dhh1, and Pab1 on stress granules in Saccharomyces cerevisiae. PLoS One *5*, e10006.

Thomas, M.G., Loschi, M., Desbats, M.A., and Boccaccio, G.L. (2011). RNA granules: the good, the bad and the ugly. Cell Signal *23*, 324-334.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of molecular biology *337*, 635-645.

6. Conclusiones

Esta tesis estudia diferentes complejos ribonucleoprotéicos del ciclo celular del ARN mensajero desde el punto de vista de sus estructuras e interacciones. Todas las proteínas caracterizadas pertenecen al microorganismo *S. cerevisiae* y se caracterizan por contener varios dominios de unión de ARN (RRMs o dedos de zinc) que presentan variaciones de comportamiento respecto del canónico: motivos estructurales adicionales, nuevas superficies de interacción y capacidad de reconocimiento proteína-proteína. En la mayoría de los casos estas nuevas características resultan claves para el desarrollo de las funciones exclusivas de cada proteína en el contexto biológico en el que se ha especializado. Del presente trabajo se pueden extraer las siguientes conclusiones:

- Se ha realizado un exhaustivo estudio químico-físico de las proteínas Gbp2p y Hrb1p y se ha determinado la estructura tridimensional por resonancia magnética nuclear (RMN) de cuatro dominios RRM ("RNA Recognition Motif") de las proteínas Gbp2p (RRM3) y Hrb1p (RRM1, RRM2 y RRM3). Encontrando que los dominios situados en la parte C-terminal de ambas proteínas presentan un plegamiento novedoso incorporando una hélice α adicional en la parte N-terminal. Los otros dos dominios de Hrb1p (y probablemente los de Gbp2p) se estructuran siguiendo una arquitectura canónica de tipo $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$.

- Se ha estudiado la interacción de Gbp2p y Hrb1p con diferentes sondas de ácidos nucleicos por RMN y otros métodos. Se ha determinado que las zonas responsables de la interacción se encuentran en los dos primeros dominios RRM para ambas proteínas, mientras que los dominios RRM3 no intervienen. Las plataformas de interacción con ácidos nucleicos involucran la zona canónica de los RRM para el caso de los RRM1 (la cara expuesta de la lámina β) y una interfase novedosa para los RRM2 compuesta por la hélice α_1 y la hebra β_2 y que aparece conservada en los dominios RRM2 de SRSF1 humana (y otras proteínas SR). Por otra parte se ha estudiado la afinidad y selectividad por diferentes sondas, hallando que estas proteínas unen preferentemente moléculas de cadena sencilla (no cuádruplex) de ARN(ADN) ricos en guaninas y uracilos (timinas), probablemente conteniendo el motivo GGU.

- Pese a su similitud con las proteínas SR implicadas en "splicing" en mamíferos, se ha descartado la función directa de Gbp2p y Hrb1p en la maquinaria del "splicing" de ARNm. En general la eliminación de ambas proteínas en levadura produce un ligero pero significativo aumento de la producción de proteínas (β-galactosidasa), probablemente correlacionado con una mayor cantidad de ARNm. El mecanismo por el cual se produce este efecto es aún desconocido.

- Se ha encontrado una función para los dominios RRM3 de Gbp2p y Hrb1p como plataformas de interacción con el complejo THO y se ha propuesto una posible superficie de interacción que se encuentra conservada en los dominios RRM1 de SRSF1 y SRSF9 humanas. La función de reconocimiento de THO es exclusiva del dominio RRM3 de ambas proteínas y se puede transferir a otras (Npl3p) mediante la construcción de quimeras.

- Se ha caracterizado la afinidad y selectividad por ARN de la región Zf1-4 de la proteína Nab2p usando la espectroscopia de anisotropía de fluorescencia; se ha definido una moderada selectividad por secuencias ricas en adeninas con una constante de disociación entorno a 2 μM, unas 50 veces mayor que la afinidad descrita para los dominios Nab2p Zf5-7 en la bibliografía.

- Mediante la caracterización de diversos mutantes, se han identificado las zonas de interacción con ARN en la región Zf1-4 de Nab2p, que involucran un reconocimiento con una gran componente electrostática del tándem Zf1-2 a través de la hélice α terminal y un reconocimiento específico de base a través de los residuos aromáticos expuestos del Zf3 del tándem Zf3-4. Los resultado se han integrado con los existentes en la bibliografía para la región Zf5-7 y se ha propuesto un modelo de reconocimiento global de toda la región de dedos de zinc de Nab2p.

- Se ha caracterizado mediante RMN la proteína Tif4631p, encontrando que se halla intrínsecamente desplegada en su parte N-terminal (1-402). Sin embargo se han identificado algunos segmentos con posibles estructuras residuales, algunos de los cuales como la secuencia Box 1 muestran tendencia a la agregación y a la formación de hidrogeles de Tif4631p.

- Se ha estudiado la interacción entre la proteínas Pub1p y Tif4631p encontrando que la zona de interacción en Pub1p reside en el dominio C-terminal entre la hélice α_1, los loops 2 y 4 y la hebra β_2. Por parte de Tif4631p se encuentran dos zonas con alta homología en las que se observa interacción con Pub1p: Péptido 1 (37-51) y Box 1 (92-108).

- Se ha descubierto que la interacción entre Pub1p y Tif4631p desencadena la agregación de ambas proteínas, dirigida por la secuencia Box1. Esta agregación es dependiente de la presencia de otras regiones de Tif4631p y podría tener implicaciones importantes en la formación de gránulos de estrés en levadura.

7. Appendixes

7.1. LIST OF OLIGONUCLEOTIDES USED IN CLONING

Oligonucleotides used in bacterial and yeast cloning grouped for each type of cloning.

	Oligo name	Sequence 5'-3'
Gbp2p bacterial cloning	GBP2_107_Fw	GCGCGGATCCAGAGACTTAGAAAGGCAATTTGACGC
	GBP2_Rv	CGCTCGAGATAAAGTACACAGGTCATGGTTCGGTTG
	GBP2_317_Fw	GGATGCCGGATCCCACATTGATGAAACTGCAGC
	GBP2_329_Fw	CGCTCGAGATTAATTATCCTGTCTTACCATTAGTT
	GBP2_197_Rv	GCGCGGATCCGGGTTTGAAGTGTTCATCATCAATT
	GBP2_287_Fw	CGCTCGAGATTATCTGGTATCTTCAAGGTCCTCACGCC
	GBP2_316_Rv	GGATGCCGGATCCCACATTGATGAAACTGCAGC
Hrb1p bacterial cloning	HRB1_144_Fw	TCCTGGATCCGCACGTGAATTGGATAGTACATACGA
	HRB1_Rv	CGCGCTCGAGCGTACTTTACGTAAGCCTTCTGAGGC
	HRB1_216_Fw	CGCGGATCCGTTAACGAAGAAGCTCGAAAATTTACCGAA
	HRB1_236_Rv	CGCGCTCGAGTTAATTATCTTGCCTGACGAAGAT
	HRB1_262_Fw	TCCTGGATCCGAAGTTATTGTAAAAAATTTACCTG
	HRB1_358_Rv	CGCGCTCGAGTTAGTTAACGGGAGAATCATCC
Gbp2p yeast cloning	TAP_Trp1_Gpb2_Fw	TGGTGGTTGTAGTTTACAGATCTCTTATGCTAGACGTGATTCCATGGAAAAGAGAAG
	TAP_Trp1_Gpb2ΔRRM3_Fw	AGGTACTGAACCAGGTCTTGCGCAGGATGCCGCTGTCCACTCCATGGAAAAGAGAAG
	TAP_Trp1_Gpb2_Rv	CATAAAGTACACAGGTCATGGTTCGGTTGGTGCTTAGGAATACGACTCACTATAGGG
	Del_His3_Gpb2_Fw	GCGAAAAGGAAACAAACATCAGCTGGATTTTTTCGCCAAGTCCATGGAAAAGAGAAG
	Del_His3_Gpb2_Rv	CATAAAGTACACAGGTCATGGTTCGGTTGGTGCTTAGGAATACGACTCACTATAGGG
	Nter-His Fw	CTTAGAAAGGCAATTTGACGCGACCAAGAGAAATTTTGAATCCATGGAAAAGAGAAG
	3'UTR-Marc RV	CATAAAGTACACAGGTCATGGTTCGGTTGGTGCTTAGGAATACGACTCACTATAGGG
	Nter-RRM3 Fw	CTTAGAAAGGCAATTTGACGCGACCAAGAGAAATTTTGAAAACAAGAGAAAGAACAATGATCG
	Gbp2_control_Fw	ACCGGAAACTAATGGTAAGACAGG
	Gbp2_control_Rv	GACAATAGCACAACCCAGAGGAG
	Del_Gbp2_control_Fw	GAAGTGGACAAAGAGACGGAAGTGA
Hrb1p yeast cloning	Del_His3_Hrb1_Fw	GAATGAGCAGATAGAACGCCAAATATTAAGTTAACAGGATTCCATGGAAAAGAGAAG
	Del_His3_Hrb1_Rv	AGATCCAATAGGTGAGAAAGTATATAGATCGAGAGTAGTTTACGACTCACTATAGGG
	Del_Hrb1_control_Fw	CGAAACACCCAGTCATAACAGCC
	Del_Hrb1_control_Rv	CAGTCTTCTTCAAAGTCTTCG
Npl3p yeast cloning	TAP_Npl3_Trp1_Fw	CAGAACCAGAGATGCTCCACGTGAAAGATCACCAACCAGGTCCATGGAAAAGAGAAG
	TAP_Npl3_RRM3_Fw	CAGAACCAGAGATGCTCCACGTGAAAGATCACCAACCAGGCACATTGATGAAACTGCAGC
	TAP_Npl3_Trp1_Rv	TGTTAATTTCTCCTTTTTTTTTCTCAACTATATAAATGGCTACGACTCACTATAGGG
Tho2p yeast cloning	TAP_Trp1_Tho2_Fw	AGGTCCCAAGGGTGGGAATTACGTCAGTAGGTACCAGAGGTCCATGGAAAAGAGAAG
	TAP_Trp1_Tho2_Rv	ACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGTAATACGACTCACTATAGGG
	Tho2_control_Fw	GCTTCATCTTACAAGCGCG
	Tho2_control_Rv	CGAGCATTAGCTGGACC

	Oligo name	Sequence 5'-3'
Hpr1p yeast cloning	TAP_Trp1_Hpr1_Fw	TTCGAACATTTCTAATGGTTCATCTACCCAAGATATGAAATCCATGGAAAAGAGAAG
	TAP_Trp1_Hpr1_Rv	GAATTTCTTATCAGTTTAAAATTTCTATTAAGAGGATAATTACGACTCACTATAGGG
	Hpr1_control_Fw	GCAAGATCCTGACAGTGGTGTTGCC
	Hpr1_control_Rv	CTGATAGGATCTCTGTGGTACGC
Mft1p yeast cloning	TAP_Trp1_Mft1_Fw	AAGCGATTTTAGTGCGTCTTCCTCTGTTGAAGAAGTAAAATCCATGGAAAAGAGAAG
	TAP_Trp1_Mft1_Rv	ATGCCTTTTCTATTTAGTAAGAGCTATGCATTATACGTGGTACGACTCACTATAGGG
	Mft1_control_Fw	CGCAGAGGAGGCAGATAGCG
	Mft1_control_Rv	ACGCACCATACAGAACTTGC
Thp2p yeast cloning	TAP_Trp1_Thp2_Fw	AAGCTATCCTGTAGATAAAGAAGGTGACATAGTTTTAGAATCCATGGAAAAGAGAAG
	TAP_Trp1_Thp2_Rv	ACTCTATCTAACTTGTGCAGGCTGGTTAAATAAAATGTGCTACGACTCACTATAGGG
	Thp2_control_Fw	CTTATGTTCGGTGATGATGC
	Thp2_control_Rv	TTTTCTCTTTCTTCCCTCCC
Mutant Nab2p Zf1-4 bacterial cloning	NAB2_R263A_FW *	GAGGGGCGTTGCGCATTGTTTCCTCACTG
	NAB2_H276A_FW *	GGTAGATCATGCCCAGCTGCACACCCAACTAAGG
	NAB2_F298A_FW *	CCGGAACTTGTGAGGCTTTACATCCAAATGAAG
	NAB2_K341AF342A_FW *	GGTATCGTTCTGTGTGCAGCTGGGGCTCTGTGTTCC
	NAB2_N348A_FW *	GGGCTCTGTGTTCCGCTCCATCATGCCCA
	NAB2_F353A_FW *	CCATCATGCCCAGCTGGTCATCCAACACC
	NAB2_N284RE285F_FW *	CCAACTAAGGTATGTAGATTCTATCCAAATTGTCC
	NAB2_D372KK373F_FW *	GATCTAATGTGGTGTAAGTTCAATTTGACATGTGATAATC
	NAB2_K384F_FW *	CCTGAGTGTAGATTTGCCCACTCTTCATTG
	NAB2_K309A_FW *	GATGAAGAGTTGATGGCGGAAATGGAAAGAAC
	NAB2_320-322_FW *	CGTGAAGAATTTCAAGCAGCAGCAGCTGATTTATTGGC
Pub1p cloning	PUB1_FA_FA_mut_FW	CCGTTGTTCCAGAACGCCGGTGCTATTTTAGATTTCAAGC
	PUB1_FA_FA_mut_RV	GCTTGAAATCTAAAATAGCACCGGCGTTCTGGAACAACGG
Tif4631p cloning	TIF4631_1_FW	CGCGGATCCATGACAGCAGAAACTGCTCACCCGACACAATCTGC
	TIF4631_82_stop_RV	CGCGCTCAGATTATCTGAAACTACCGCCACC
	TIF4631_186_stop_RV	GCGCTCGAGTTAAGGGGTAGGAGTTGGAGTAGAAG
	TIF4631_229_STBAM_FW	GAGAAAGGAGCAACTTTAAGGATCCAGTGGCAACAATAATATTCC
	TIF4631_229_STBAM_RV	GGAATATTATTGTTGCCACTGGATCCTTAAAGTTGCTCCTTTCTC
	TIF4631_305_stop_RV	GCGCTCGAGTTACTTCAAACGTTCAGCAAAGGTTAAC
	TIF4631_348_stop_RV	GCGCTCGAGTTATTTAACCTGTTCACTGGGAGGC
	TIF4631_394_stop_RV	CGCGCTCGAGTTAATCAGTTGTAGTTTCGATTTCAGCTTCAAGTCC
	TIF4631_W95_FW	GCTCAAACGTGCCAGCGACTGGTTACTATAATAAC
	TIF4631_W95_RV	GTTATTATAGTAACCAGTCGCTGGCACGTTTGAGC
	TIF4631_F98F99_FW	GTGCCATGGACTGGTGCCGCTAATAACTACCCCG
	TIF4631_F98F99_RV	CGGGGTAGTTATTAGCGGCACCAGTCCATGGCAC
	TIF4631_F105F106_FW	CTATAATAACTACCCCGTTGCCGCCCAGCCCCAGCAA
	TIF4631_F105F106_FW	TTGCTGGGGCTGGGCGGCAACGGGGTAGTTATTATAG

Table 7.1. – List of DNA oligonucleotides used as primers in different PCRs. * Means that reverse complementary oligo was also used.

7.2. HRB1p AND GBP2p SECONDARY STRUCTURE PREDICTION

Several web sources were used for different predictions based on sequence comparison. First, DISOPRED (Ward, McGuffin et al. 2004), an unstructured region detector, was used in a preliminary analysis, finding the three RRM domains as the only structured regions (Figure 7.1).

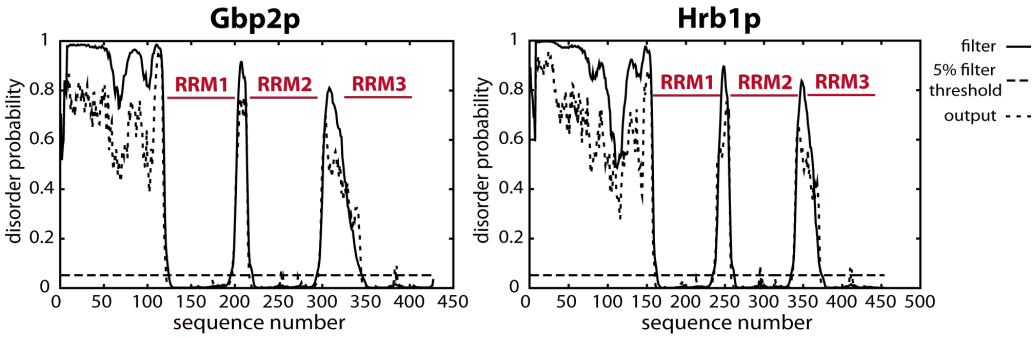


Figure 7.1. –Unstructured sequence prediction of Gbp2p and Hrb1p proteins made by DISOPRED programme. The structured zones (gaps) identify the proposed domain regions with high precision.

For secondary structure elements (α-helixes and β-strands) recognition, the following set of programs were used: PSIPRED from London University College (http://bioinf.cs.ucl.ac.uk/psipred/), ss_Predictions program from CFSSP server (http://www.biogem.org/tool/chou-fasman/), the advanced protein secondary structure prediction server APSSP2 (http://imtech.res.in/raghava/apssp/), PROF server (named as MODEL) (http://www.aber.ac.uk/~phiwww/prof/) and jnetpred, JNETHMM and JNETPSSM, which are different predictions incorporated in the Jpred3 program (http://www.compbio.dundee.ac.uk/www-jpred/). In addition, two other programs were used for coiled coil motifs detection: coiled coil predictor Parcoil2 (http://groups.csail.mit.edu/cb/paircoil2/) and Lupas 14, this one incorporated in Jpred3. All results are summarised in the Figure 7.2.

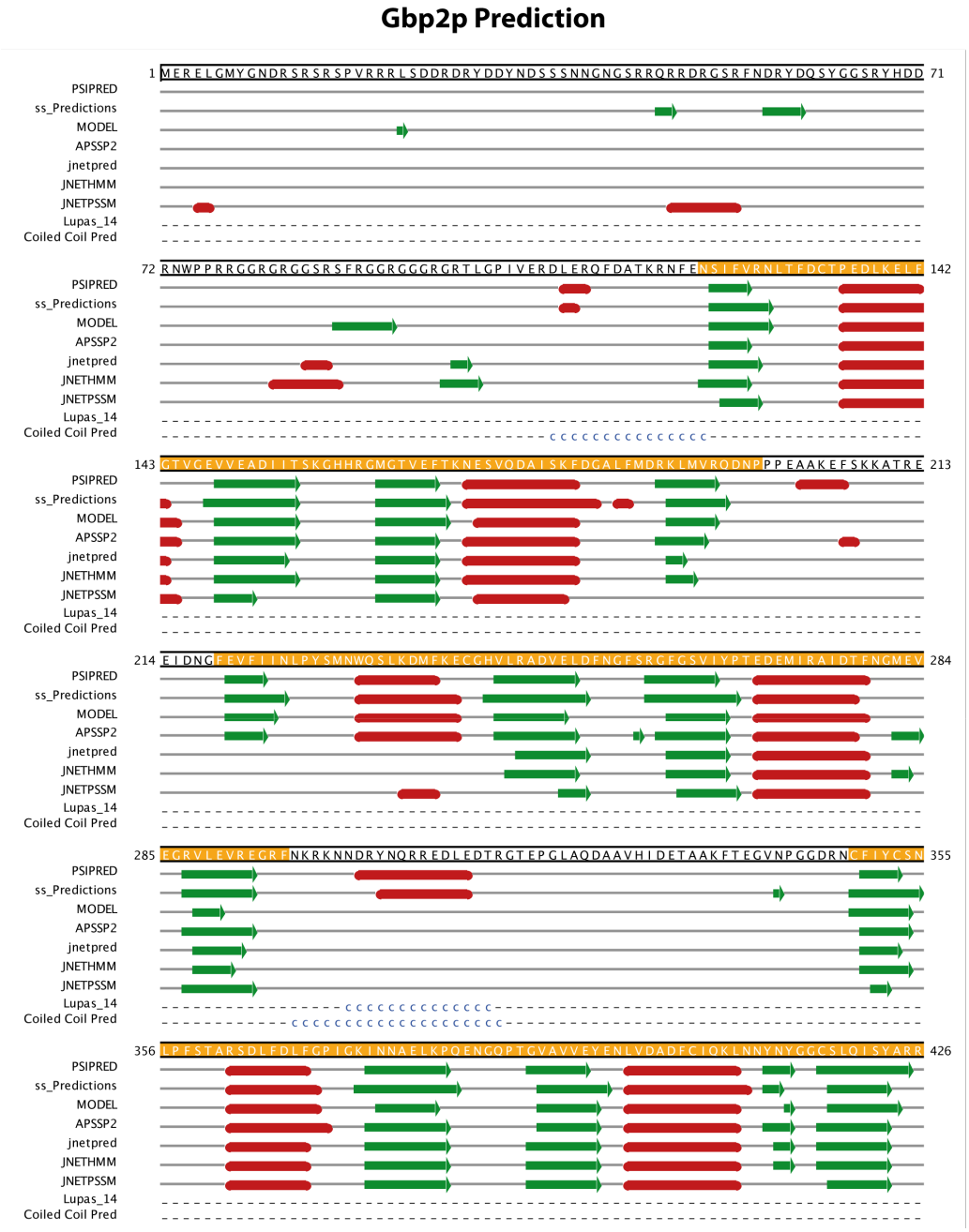
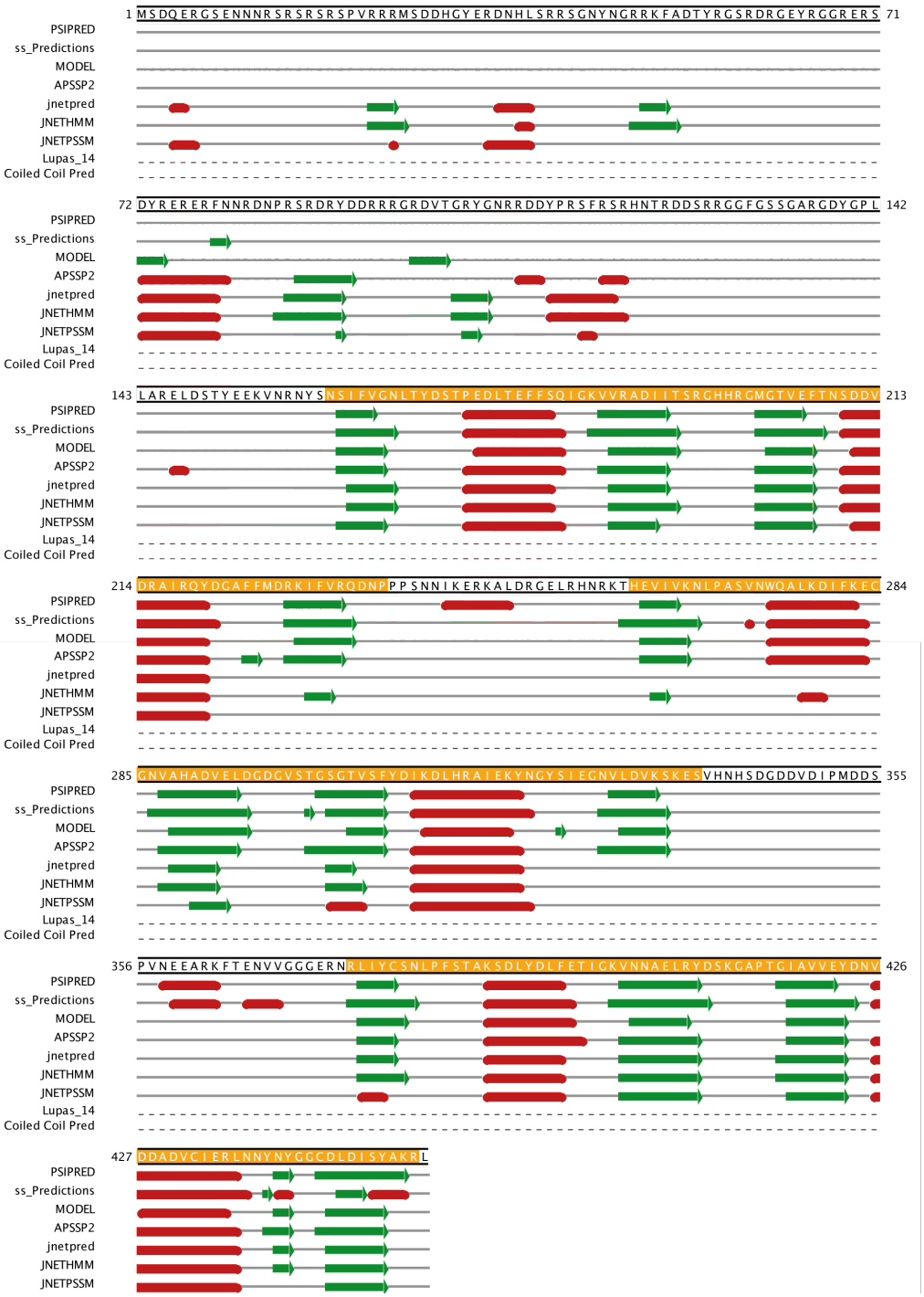


Figure 7.2. – Secondary structure prediction summary of Gbp2p and Hrb1p obtained from several web-based resources. α-helixes are represented by red tubes, β-strands by green arrows and the letter c marks coiled coil regions. The proposed RRMs domains are outlined in orange in the sequence, finally the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ arrangement is appropriately predicted for all of them.

Hrb1 Prediction



7.3. ALIGNMENT OF GBP2p AND HRB1p HOMOLOGOUS PROTEINS

The alignment (Figure 7.3) was made using the detected Gbp2p and Hrb1p homologous by fungal BLAST (http://www.yeastgenome.org/cgi-bin/blast-fungal.pl), excluding those homologous proteins from *Saccharomyces* genus.

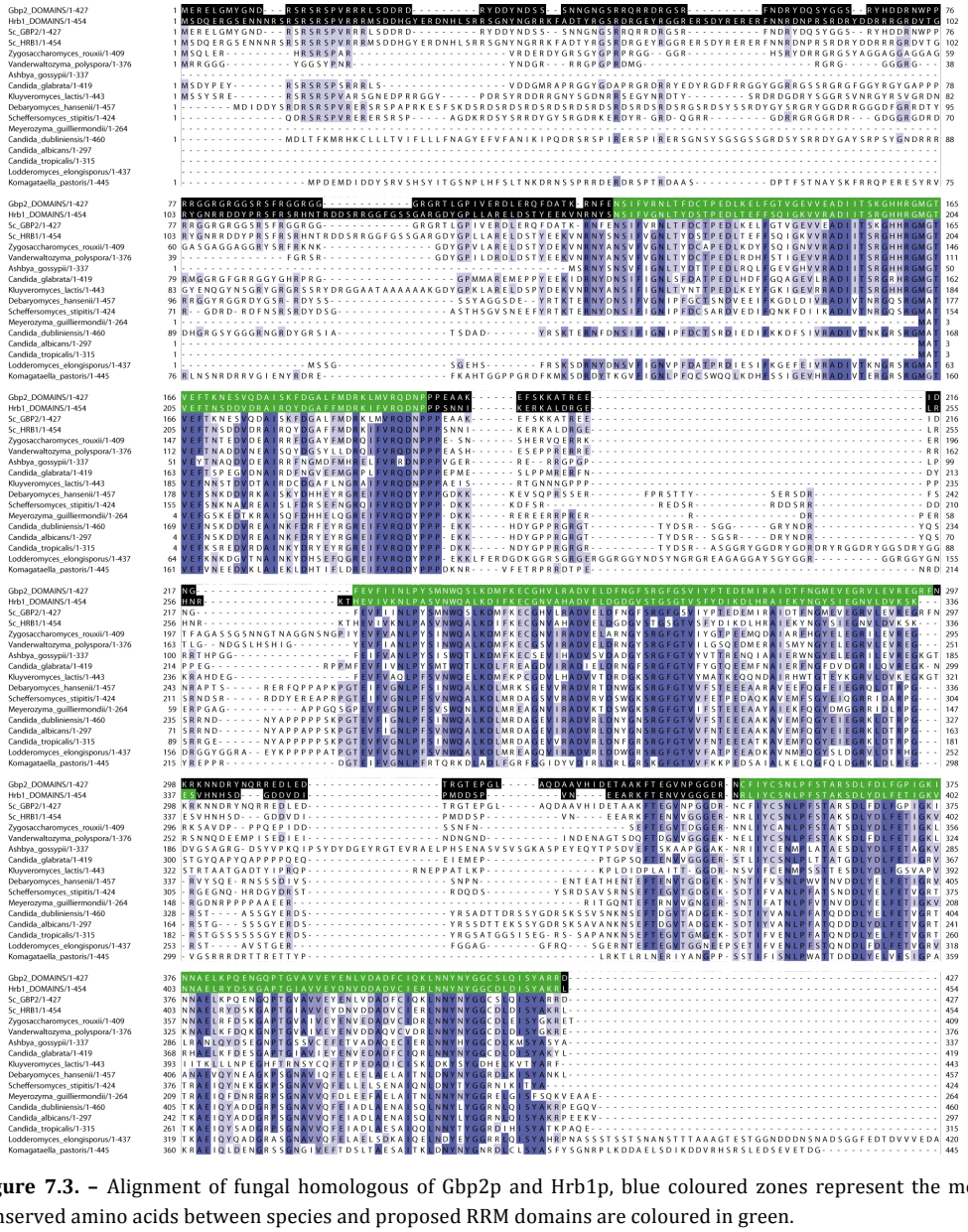


Figure 7.3. – Alignment of fungal homologous of Gbp2p and Hrb1p, blue coloured zones represent the most conserved amino acids between species and proposed RRM domains are coloured in green.

7.4. CHEMICAL STABILITY OF HRB1p AND GBP2p CONSTRUCTS

Initial NMR data of Hrb1p and Gbp2p constructs showed sample heterogeneity. A comprehensive analysis using PAGE-SDS, gel filtration, NMR and mass spectrometry was made to characterize the origins of this heterogeneity, concluding that cysteine reactivity is the most likely source of these effects. Based on this study the purification protocols could be improved to avoid protein heterogeneity. Also, some practical conclusions of general interest about protein purification could be outlined.

7.4.1. Disulphide bridge oligomerization

Some constructs of Hrb1p and Gbp2p proteins presented several oligomeric forms detectable in PAGE and gel filtration chromatography experiments, which depended on the reduction conditions used along purification protocols. Oligomerization was clearly observed in Hrb1p RRM23, Hrb1p RRM123 and Gbp2p RRM123 constructs when proteins were purified in buffers with low concentration of reduction agent (Figure 7.4.A.) and disappeared at higher concentration of reduction agents.

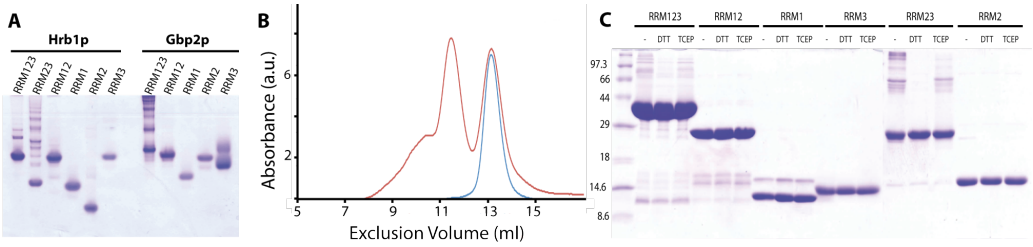


Figure 7.4. – A) Native PAGE of all Gbp2p and Hrb1p constructs. Oligomer can be seen for Hrb1p RRM123, Hrb1p RRM23 and in Gbp2p RRM123 constructs. **B)** Gel filtration chromatography of Gbp2p RRM123 construct purified using buffers with high (1 mM) (blue) and low (0 mM) (red) concentration of reduction agents (DTT). Mobile phase used during the chromatography (Superdex 200) contains low concentration of reduction agent (0,1 mM of DTT) **C)** Compared efficiency of different reduction agents for Hrb1p constructs. Samples for each construct where incubated with no agents (first line of each), 5 mM DTT (second line) and 1 mM TCEP.

Oligomer formation seems to happen during chromatography steps. Once purified, the monomers remain stable even upon decrease of the concentration of reduction agents (Figure 7.4.B.). Additional analysis showed that this protein oligomerization could be reverted under strong reduction conditions, in particular with DTT (Figure 7.4.C.).

All these data suggest an oligomerization process dependent on intermolecular disulphide bridge formation. Moreover, this oligomerization seems to be strongly favoured during purification procedure likely due to the increment of the local protein concentration inside the chromatographic columns. The use of strong reduction buffers during purification steps is critical to avoid this process. Once proteins are purified, the concentration of reduction agent could be lowered to minimize the interference that these chemicals might cause in the biophysical study of these proteins.

7.4.2. Cysteine chemical modification

Disulphide bridge oligomerization in not the only source of cysteine-mediated sample heterogeneity. A more rare effect involving this type of residues was observed in many samples of both proteins.

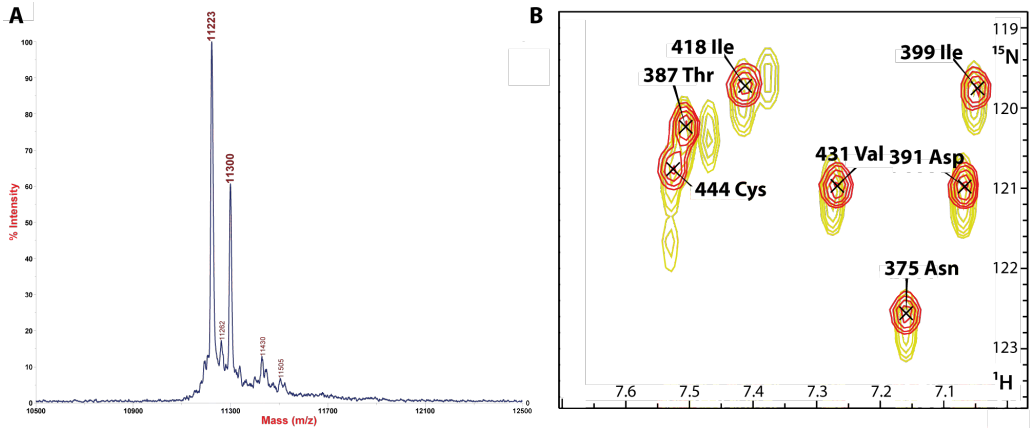


Figure 7.5. – A) Mass spectrometry chromatogram of Hrb1p RRM3 partially modified protein, where two main peaks appear with a difference of 77 Da between them. **B)** ^{1}H-^{15}N HSQC spectra section superposition of two different Hrb1p RRM3 proteins, in yellow it is shown the spectra of the partially modified protein (similar as in figure A) and in red the spectra of the non modified protein.

NMR and mass spectrometry data showed the presence of modified species in several constructs (Gbp2p RRM123, RRM3, Hrb1p RRM123, RRM2 and RRM3). The major specie was 75 mass units heavier than the wild type one (Figure 7.5) and both were detected coexisting in the same samples with different degrees of modification (reaching in some cases 100% of modified protein).

Duplicated residue resonances were detected in NMR spectra and they belonged to RRM3 domains of both proteins and to Hrb1p RRM2. Moreover, these residues mapped in clusters (loops 1, 3 and 5) on structures close to an exposed cysteine (Figure 7.6) that was suspected to be at the origin of the effect.

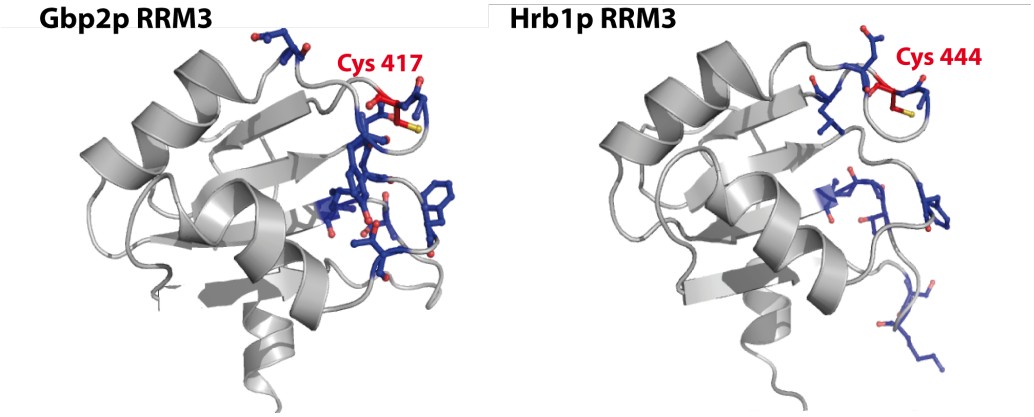


Figure 7.6. – Representation of Gbp2p and Hrb1p RRM3 domains, colouring in blue the residues affected by the modification and in red the exposed cysteines, also shifted in NMR spectra.

To identify the chemical nature of this modification, many tests were made changing purification conditions. From the analysis of the results obtained using protocols with variations in the type of reduction agent, its concentration and in the buffer composition it could be concluded that:

- The nature and concentration of reduction agent used in lysis, nickel affinity chromatographies and TEV digestion (see materials and methods) did not alter the chemical integrity of the proteins.
- The buffer and the reduction agent used during ion exchange chromatography determine the presence of sample heterogeneity.

A TEV-digested Hrb1p RRM3 sample was divided and purified by ion exchange (monoQ) using buffers with 1 mM DTT, 1 mM β-mercaptoethanol or without reduction agent (buffer composition: 20 mM Tris-HCl pH 8.0, 25 – 1000 mM NaCl). Modification only occurred when β-mercaptoethanol was used (Figure 7.7.A.). In addition, another experiment was performed to test if the column matrix has some catalytic role in the reaction. The same unmodified sample of Hrb1 RRM3 was incubated overnight at 4ºC in three different buffers: 50 mM Tris-HCl pH 8.0, 25 mM NaCl, 10 mM DTT; 50 mM Tris-HCl pH 8.0, 25 mM NaCl, 10 mM β-

mercaptoethanol; 50 mM potassium phosphate pH 8.0, 25 mM NaCl, 10 mM β-mercaptoethanol. The results showed a modification peak in β-mercaptoethanol buffers and this peak was bigger in Tris-HCl buffer than in phosphate one (Figure 7.7.B.). Although the modification takes place in these conditions, the reaction yield is much lower than in the first case (when protein was loaded into the column) even though β-mercaptoethanol concentration is 10 fold higher.

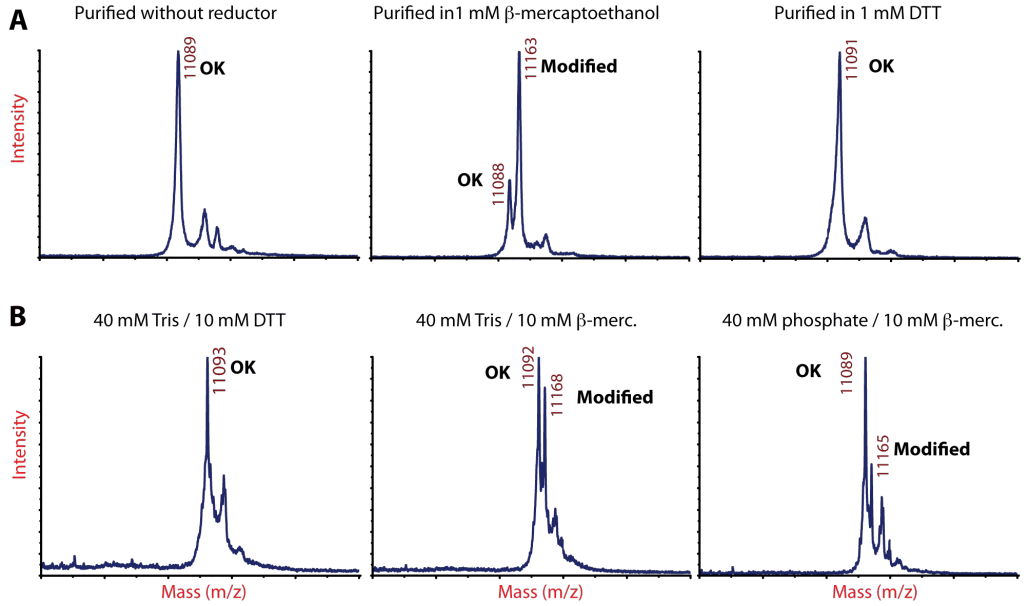


Figure 7.7. – A) Mass spectrometry spectra of three Hrb1p RRM3 samples purified in anion exchange chromatography using three different conditions: without reductor, using 1 mM β-mercaptoethanol and using 1 mM DTT. The protein purified in β-mercaptoethanol, presents two peaks, the original protein (around 11088 Da) and the modified one with 75 Da more. **B)** Mass spectrometry spectra of Hrb1 RRM samples incubated with different buffers; modified protein appeared in β-mercaptoethanol containing ones.

All these data suggest that the modification is due to the linkage of a molecule of β-mercaptoethanol to a cysteine side chain. The reaction is favoured by Tris-HCl buffers and strongly catalysed either by the ion exchange column, either by the effect of increase of protein concentration or by an unknown chemical element in the column matrix.

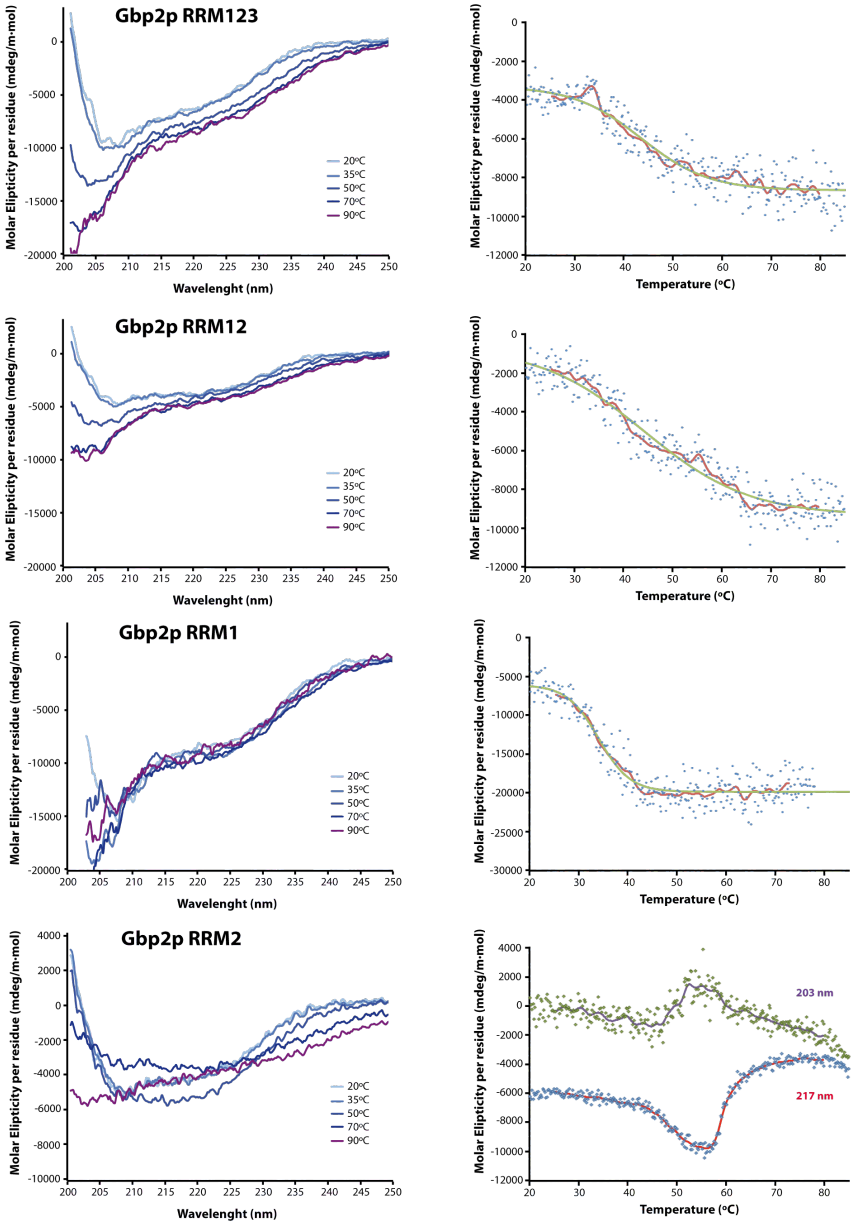
Chemical modifications of proteins by β-mercaptoethanol have been reported (Fujiwara, Nakano et al. 2007, Zhukov, Ejchart et al. 2008) and, as in the Gbp2p/Hrb1p cases, the modified proteins have molecular weights 75 units larger than natural ones. This difference is close to the molecular weight of the β-mercaptoethanol (78.13 g/mol). Recently the three

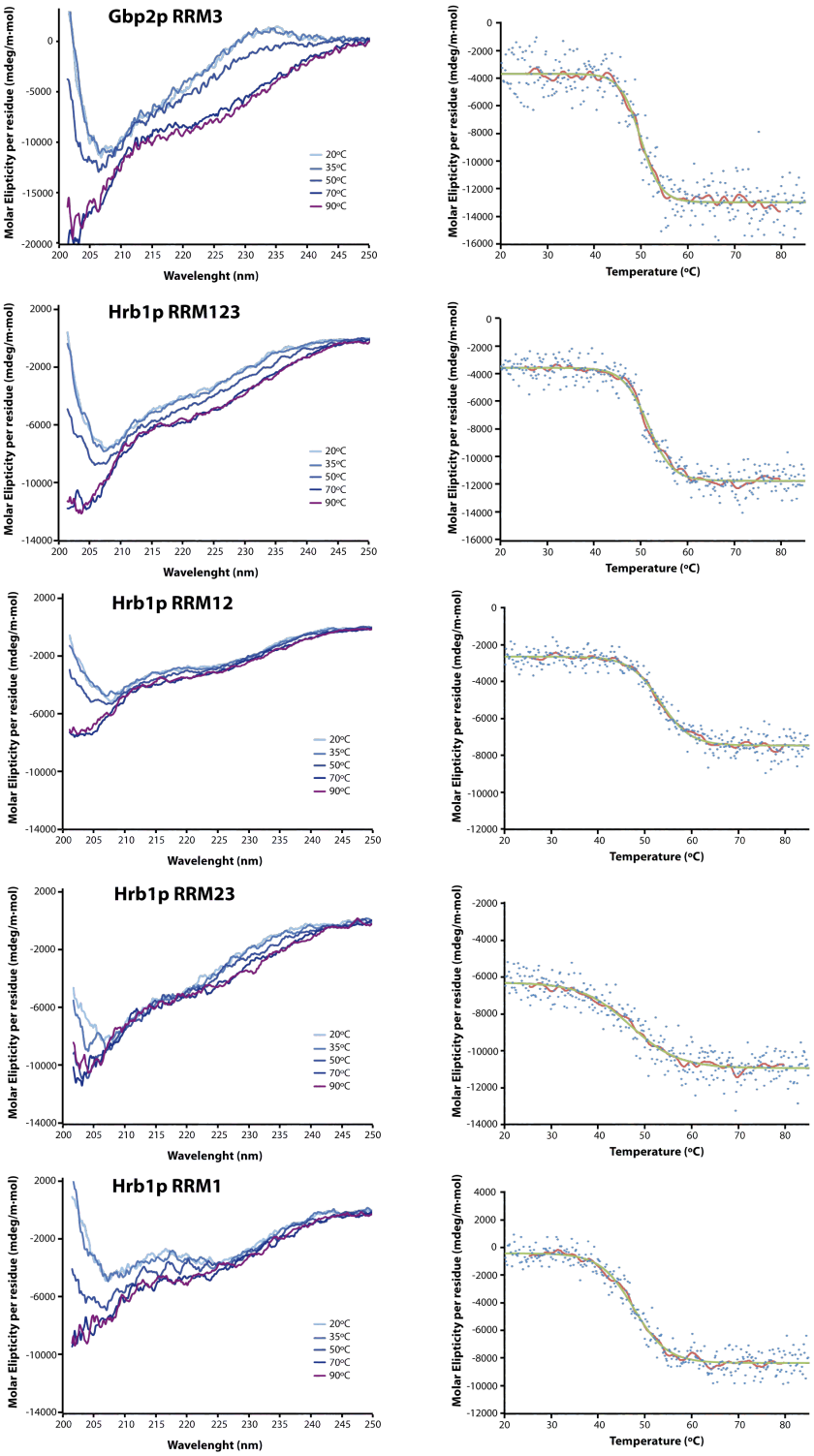
dimensional structure of the Cu,Zn SOD1 (superoxide dismutase 1) (Ihara, Fujiwara et al. 2012) showed that β-mercaptoethanol forms a disulphide bridge with the exposed cysteine, which increase protein stability substantially. Reduction of this S-S bond was attempted for Hrb1p RRM3 samples by incubating with 10 mM DTT, 13 mM β-mercaptoethanol and 1 mM TCEP respectively. The progress of the reaction was followed by NMR showing no signs of reversibility after 24 hours, indicating the extraordinarily stability of this type of modifications.

In practice, β-mercaptoethanol was not used after nickel reverse purification either in further purification steps or in different experiment conditions in order to avoid the formation of the modified species. Typically, 0.1 mM DTT was used in experimental conditions to maintain cysteines in the reduced state.

7.5. CIRCULAR DICHROISM DATA

All circular dichroism spectra obtained at different temperatures and the denaturing curves for each Gbp2p and Hrb1p construct are shown in Figure 7.8.




Gbp2p RRM3
Hrb1p RRM123
Hrb1p RRM12
Hrb1p RRM23
Hrb1p RRM1
Molar Elipticity per residue (mdeg/m·mol)
Wavelenght (nm)
Temperature (ºC)
20ºC
35ºC
50ºC
70ºC
90ºC

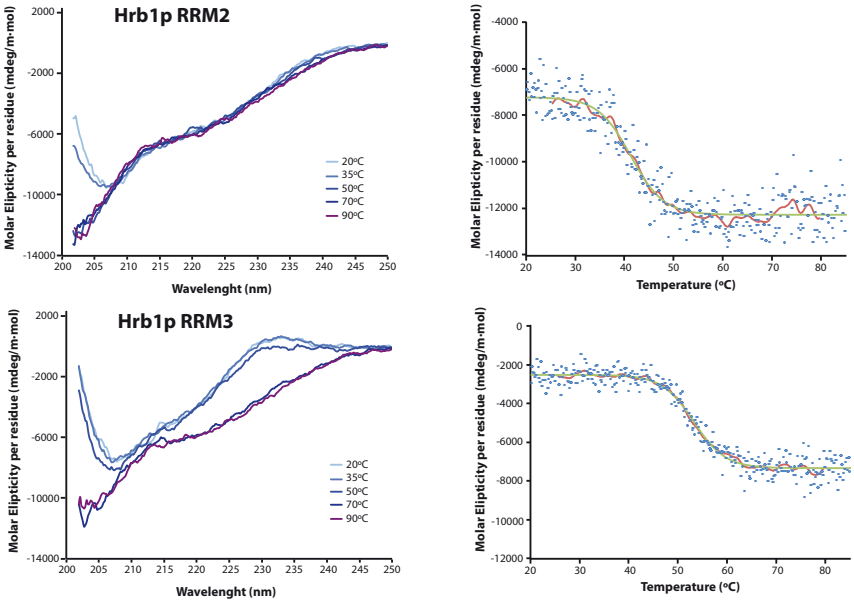


Figure 7.8. – Representation of circular dichroism spectra at different temperatures (left graph) and thermal denaturalization process (right graph; obtained data is represented in blue points, in red lines the smoothed data and in green it is drawn the sigmoidal fitting curve) of all Gbp2p and Hrb1p constructs.

7.6. CRYSTALLIZATION ASSAYS OF HRB1p AND GBP2p RRM12

The structures of the Gbp2p/Hrb1p RRM12 tandems were attempted by X-ray crystallography. All crystallization trials started from concentrated protein stocks (Gbp2p RRM12 at 19.4 mg/ml and Hrb1p RRM12 at 16.4 mg/ml) in a buffer containing 25 mM potassium phosphate and 25 mM NaCl. Exploratory condictions were set using Nanodrop Exty robot (Innovadyne Technologies Inc.) for drop mixing (250 nl of protein stock with 250 nl of each buffer) in 96-weels sitting drop plates. Six different commercial sets of conditions were used for each protein: Salt Rx, Index, Crystal Screen (Hampton Research), PACT suite (QUIAGEN), JBScreen Classic and JCSG+suite (Jena Bioscience). Plates were incubated at 18ºC during several weeks and each condition was regularly checked for crystal formation under microscope.

Clear crystals were not obtained in any case, but a set of promising conditions were selected for repeating the screening in 24 wheels hanging drop plates: for Hrb1p RRM12 construct, a screening of different phosphate salts (potassium and sodium salts, ranging pH from 5.0 to 8.2 and salt concentrations from 0.8 to 1.8 M), a citrate screening (using ammonium and sodium salts in different pH conditions form 4.8 to 8.0 and ranging concentrations from 0.8 to 2.0 M) and a malonate screening (ranging pH from 4 to 8 and concentrations from 1 to 2.5 M) were carried out and for Gbp2p RRM12 construct a screening of different PEG 4000 solutions was made (in different pH buffers from 4.0 to 6.5 and using concentrations ranging from 5 to 30 %). As in the first screening, plates were incubated at 18ºC and crystal formation was checked regularly, but no crystals appeared.

As a final attempt, a buffer was selected for each protein (for Gbp2p RRM12 15% PEG 4000 at pH 4.6 using sodium acetate; for Hrb1p RRM12 1.75 M sodium citrate at pH 5.5) to perform an additive screening. Using the robot for drop mixing and plating, a commercial additive set of conditions (Additive Screening HR2-428, Hampton Research) was tested. Plates were stored and regularly checked as in other cases, obtaining the same results.

One explanation of the unsuccessful crystal formation could be that both RRM domains are quite independent structures, thus regular crystallization is quite difficult. An alternative approach to try is to crystallize both constructs in the presence of DNA or RNA; both domains should bind to a same oligonucleotide making the complex more rigid and promoting regular crystallization.

7.7. DEAMIDATION OF TIF4631p IDENTIFIED BY NMR

Deamidation is a chemical modification of peptides in which an asparagine residue is transformed into an isoaspartate one through a cyclic intermediate (succinimide ring) (Figure 7.9). This reaction consists on the nucleophilic attack of the peptide bond nitrogen atom of the Asn following residue on the side chain carbonyl group, resulting in the deamidation of the asparagine. The succinimidyl intermediate then undergoes a hydrolysis at two different positions generating aspartic acid or isoaspartate (in a ratio 1:3).

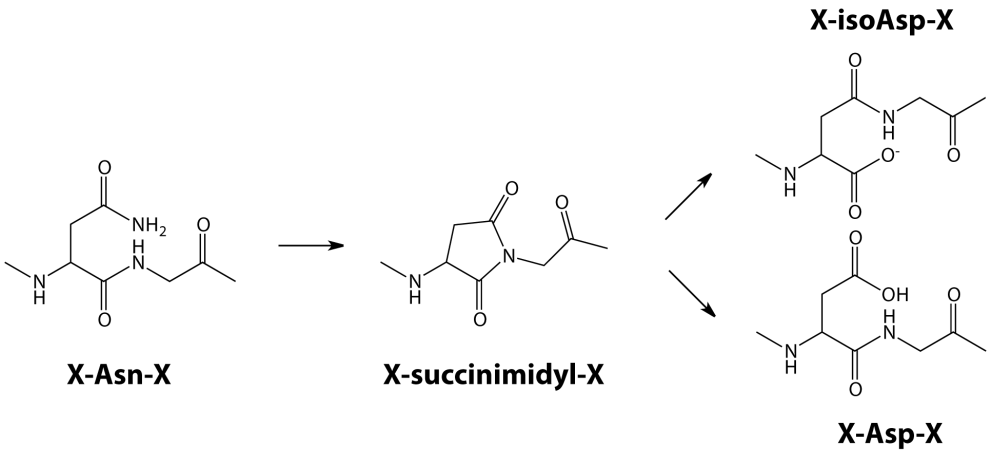


Figure 7.9. – Deamidation of asparagine residue process.

This process occurs spontaneously under physiological conditions and has been reported in many proteins, both *in vitro* (recombinant proteins) and *in vivo*. Moreover, cells have a repair system to reverse this aging effect that could play an important role in accelerating the amyloid formation in brain tissues (Dunkelberger, Buchanan et al. 2012).

Deamidation rates depend on the type of residue that follows the asparagine (G>S>>N) (Robinson 2002), but also on the three-dimensional structure of the protein, the temperature and pH conditions, the buffer composition, etc. (Robinson and Robinson 2001).

In the case of Tif4631p, deamidation was detected in two asparagine positions (Asn 41 and Asn 76) and in both cases followed by a glycine residue. Their assignment was possible using the standard backbone assignment NMR experiments, in which the typical nuclei connectivity (see 2.13.1.2 section) was altered due to the chemical modification. ^{1}H-^{15}N HSQC spectra of long aged samples presented some chemical shift perturbations and apparition of new peaks

when were compared with fresh ones (Figure 5.10.C). And following the assignment route, the isoaspartate nuclei could be assigned (Figure 7.10).

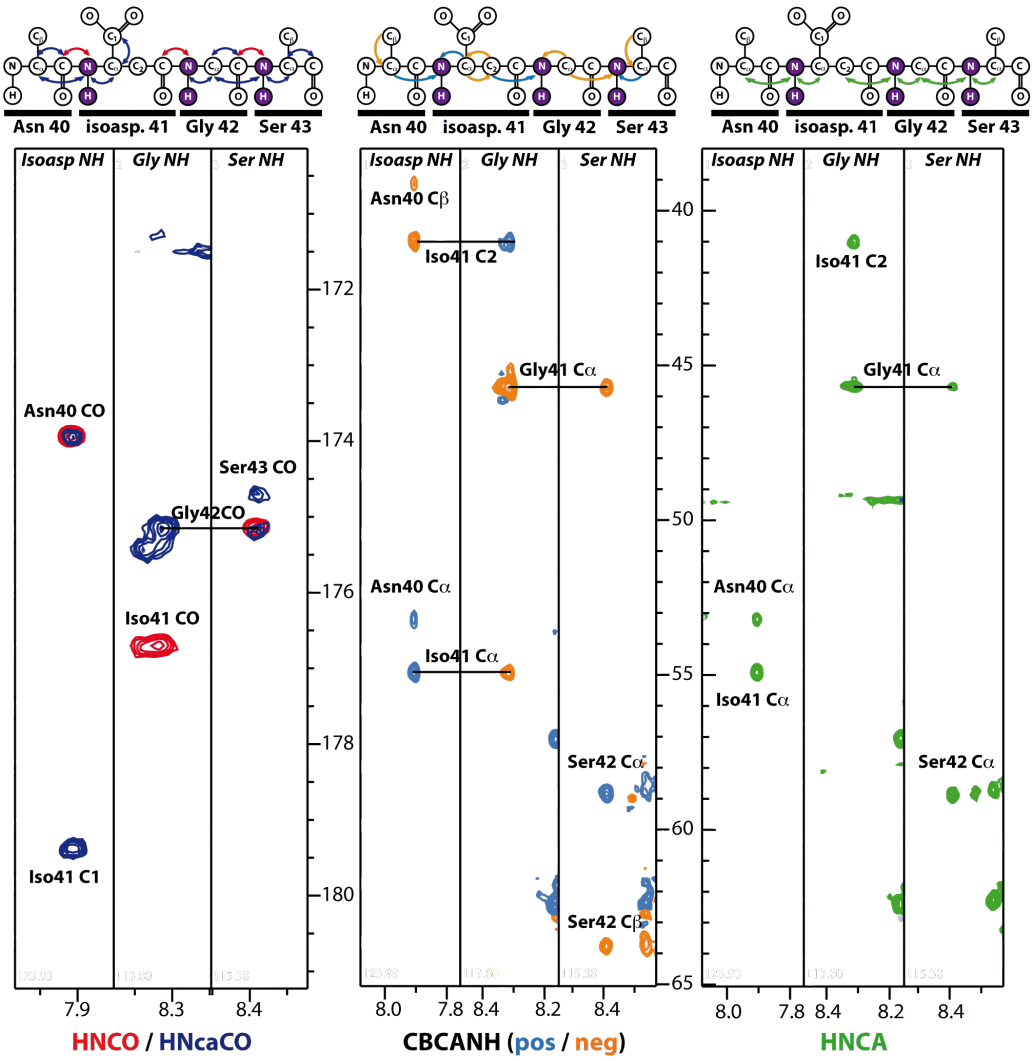


Figure 7.10. – Strips of different three-dimensional NMR spectra showing the connectivity between the residues surrounding the isoaspartate residue.

7.8. REFERENCES

Dunkelberger, E. B., L. E. Buchanan, P. Marek, P. Cao, D. P. Raleigh and M. T. Zanni (2012). "Deamidation accelerates amyloid formation and alters amylin fiber structure." J Am Chem Soc **134**(30): 12658-12667.

Fujiwara, N., M. Nakano, S. Kato, D. Yoshihara, T. Ookawara, H. Eguchi, N. Taniguchi and K. Suzuki (2007). "Oxidative modification to cysteine sulfonic acid of Cys111 in human copper-zinc superoxide dismutase." J Biol Chem **282**(49): 35933-35944.

Ihara, K., N. Fujiwara, Y. Yamaguchi, H. Torigoe, S. Wakatsuki, N. Taniguchi and K. Suzuki (2012). "Structural switching of Cu,Zn-superoxide dismutases at loop VI: insights from the crystal structure of 2-mercaptoethanol-modified enzyme." Biosci Rep **32**(6): 539-548.

Robinson, N. E. (2002). "Protein deamidation." Proc Natl Acad Sci U S A **99**(8): 5283-5288.

Robinson, N. E. and A. B. Robinson (2001). "Prediction of protein deamidation rates from primary and three-dimensional structure." Proc Natl Acad Sci U S A **98**(8): 4367-4372.

Ward, J. J., L. J. McGuffin, K. Bryson, B. F. Buxton and D. T. Jones (2004). "The DISOPRED server for the prediction of protein disorder." Bioinformatics **20**(13): 2138-2139.

Zhukov, I., A. Ejchart and A. Bierzynski (2008). "Structural and motional changes induced in apo-S100A1 protein by the disulfide formation between its Cys 85 residue and beta-mercaptoethanol." Biochemistry **47**(2): 640-650.

7.9. ABBREVIATIONS

Abs Absorbance

AcO Acetate

CD Circular Dichroism

cDNA Complementary Deoxyribonucleic Acid

CSI Chemical Shift Index

CSM Complete Synthetic Medium

CPSF Cleavage and Polyadenilation Specificity Factor

DIPSI Decoupling In the Presence of Scalar Interactions

DMSO Dimethylsulfoxide

DNA Deoxyribonucleic Acid

DSS sodium 2,2-dimethyl-2-silapentane-5-sulphonate

DTT Dithiothreitol

EDTA Ethylendiamintetraacetic acid

eIF eukaryotic Initiation Factor

ε_m Molar Absorbance Coefficient

EMSA Electromobility Shift Assay

FPLC Fast Protein Liquid Chromatography

GLH Glutaraldehyde

GST Glutathione S Transferase

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HSQC Heteronuclear Simple Quantum Correlation

IgG Immunoglobulin G

INEPT Insensitive Nuclei Enhanced by Polarization Transfer

IPTG Isopropyl β-D-1-thiogalactopyranoside

IUP Intrinsically Unstructured Protein

J Coupling constant

KOD *Thermococcus kodakaraensis*

λ Wavelength

LB Lysogenic Broth

MOPS 3-(N-morpholino)propanesulfonic acid

MoRF Molecular Recognition Feature

mRNA Messenger Ribonucleic Acid

mRNP Messenger Ribonucloprotein Particle

MW Molecular Weight

OD Optical Dispersion

NMR Nuclear Magnetic Resonance

NPC Nuclear Pore Complex

NOESY Nuclear Overhauser Effect Spectroscopy

ONPG Ortho-nitrophenyl-β-galactoside

PAGE Polyacrylamide Gel Electrophoresis

PAP Peroxidase Anti-Peroxidase

PABPs Poly (A) Binding Proteins

PB Processing Bodies

PBS Phosphate Buffered Saline

PEG Polyethylenglycol

PIC Pre-Initiation Complex

PIPES piperazine-N,N′-bis(2-ethanesulfonic acid)

PMSF Phenylmethylsulfonyl fluoride

PCR Polymerase Chain Reaction

RMSD Root Mean Square Deviation

RNA Ribonucleic Acid

RRM RNA Recognition Motif

rNTP Ribonucleotide triphosphate

SG Stress Granules

SDS Sodium dodecyl sulphate

θ Ellipticity

TAP Tandem Affinity Purification

Taq *Thermus aquaticus*

TB Tris-base / Boric acid buffer

TBE Tris-base / Boric acid / EDTA buffer

TE Tris-HCl / EDTA buffer

TEMED N,N,N,N-tetramethylethylene-diamine

TEV Tobacco Etch Virus

TOCSY Total Correlation Spectroscopy

Tris 2-Amino-2-hydroxymethyl-propane-1,3-diol

TxA Thiorredoxine A

UTR Untranslated Region

UV Ultraviolet

YPD Yeast extract Peptone Dextrose medium

Zf Zinc finger